Met Office

National Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

NEMO-Surge: Forecast performance during 2016-2017 winter trial Technical report 622

August 2017

Clare O'Neill, Andrew Saulter

Contents

Summary **2**

1 Introduction **3**

2 Operational storm surge forecast system **3**

3 Analysis methods **4**

3.1 Observations 5

3.2 Qualitative analysis (case studies) 8

3.2.1 Note on storm names 9

3.3 Quantitative analysis (overall statistics) 9

4 Case studies: North Sea coast **10**

4.1 26 & 27 December 2016 (storm Conor) 10

4.2 13 January 2017 15

5 Case studies: Irish Sea coast **22**

5.1 23 December 2016 (storm Barbara) 22

5.2 23 & 26 February 2017 (storms Doris and Ewan) 25

6 Case studies: South coast **31**

6.1 13–14 January 2017 31

6.2 28–29 May 2017 34

7 Overall statistics **36**

8 Discussion and conclusions **38**

References **39**

© Crown Copyright 2017

Summary

The NEMO-Surge prototype model described by O'Neill et al. (2016) was introduced into the Met Office operational system in November 2016, and has been running successfully in parallel with the existing CS3X surge model since then.

NEMO-Surge forecasts from the period, December 2016–May 2017 have been analysed and compared with the equivalent CS3X forecasts and observations. Although this winter storm period was relatively quiet, there are a number of surge case studies to examine. These show that NEMO-Surge and CS3X produce very similar forecasts and in most cases they are almost indistinguishable. There are some examples with slightly larger differences, including some where NEMO-Surge is better compared with observations and others where CS3X is better. These differences are in the fine details of the forecast though, with the overall picture remaining the same from both models.

Quantitative statistical comparison done over the six month trial period shows almost no difference between NEMO-Surge and CS3X at all forecast leadtimes.

Overall, this first winter trial season has confirmed that NEMO-Surge is a viable replacement for CS3X. The next stages will be to support customers to receive trial data feeds for testing in their downstream processing systems ahead of the 2017-18 storm season. In the longer term, work is needed to understand the sources of bias in the models, particularly discrepancies in the constituents included in harmonic tide predictions, before further improvements to model performance can be made.

© Crown Copyright 2017

1 Introduction

The Met Office currently runs an operational storm surge forecast system for the UK. This system uses CS3X (Flather, 1994) as the underlying ocean model and runs in both deterministic and ensemble modes (Flowerdew et al., 2010). In order to harmonise the operational models run at the Met Office and facilitate future model improvements, work is now underway to assess whether a 2D (barotropic) configuration of NEMO is a viable replacement for CS3X in the storm surge forecast system.

Initial sensitivity experiments of NEMO-Surge have indicated that its performance is comparable with or better than the existing CS3X model for both tide (Furner et al., 2016) and hindcast surge (O'Neill et al., 2016) predictions. As a result a prototype configuration of NEMO-Surge was introduced into the Met Office operational system in November 2016, running in parallel with CS3X, to allow assessment of forecast performance to be made over the winter storm season. This report describes the relative performance of NEMO-Surge during this winter trial period, using CS3X as a benchmark. The period used for assessment is December 2016–May 2017, allowing a month (November 2016) for the system to fully spin up. Particular focus is given to case studies of notable surge events that occurred during this time.

The current operational forecast system is first briefly described in Section 2, and the observations and analysis methods used are described in Section 3. Case studies of notable surge events are presented for ports along the North Sea coast (section 4), Irish Sea coast (section 5) and South coast (section 6) . Some overall statistical performance is presented in section 7, with discussion and final conclusions in section 8.

2 Operational storm surge forecast system

The surge model requires input forcing files of 10 m wind speed and sea level pressure, which are currently provided by the global configuration of the Met Office Global and Regional Ensemble Prediction System (MOGREPS-G). MOGREPS-G runs 4 times per day, with cycle times 00/06/12/18Z, so the storm surge model therefore follows the same run schedule. Surge model outputs are available to forecasters approximately 7 hours after the cycle time.

MOGREPS-G runs as an ensemble with initial condition perturbations provided by Ensemble Transform Kalman Filter (ETKF). During the 2016–2017 winter period considered here, the system comprised 11 perturbed members, plus the unperturbed control member, run out to T+174 hours, with a horizontal resolution of approximately 33 km. At each cycle, these 12 members are combined with the 12 from the previous run to create a time-lagged ensemble of 24 members out to T+168 hours. Further details of the MOGREPS-G system are given by Bowler et al. (2008). At the time

© Crown Copyright 2017

of writing this system has just undergone a significant upgrade, increasing to 36 members (18 + 18 time-lagged) with horizontal resolution increased to 20 km.

The storm surge model is forced by the resulting ensemble forecasts from the atmosphere model (using hourly fields), producing an ensemble of surge forecasts. A tide-only run is also done which is subtracted from the modelled total water level, leaving the predicted surge residual. The harmonic tide prediction (which has traditionally been assumed to be more accurate than the model tide) is then added back onto this residual to give the final predicted total water level. Full details of the operational surge forecast system are given by Flowerdew et al. (2010). It should be noted that at present the surge ensemble only reflects uncertainty in the atmospheric conditions, and not in the ocean model or tidal predictions.

3 Analysis methods

A number of different parameters are used in this report to compare NEMO-Surge and CS3X:

- Total water level—Water level including both tide and surge
- Residual—Difference between total water level and the level expected from tide alone
- Skew surge—Difference in height of the expected (tide-only) peak at high tide and the total peak water level seen, ignoring any difference in the timing

These are all illustrated in Figure 1.

Note that water levels for all locations are presented relative to Ordnance Datum rather than Chart Datum.

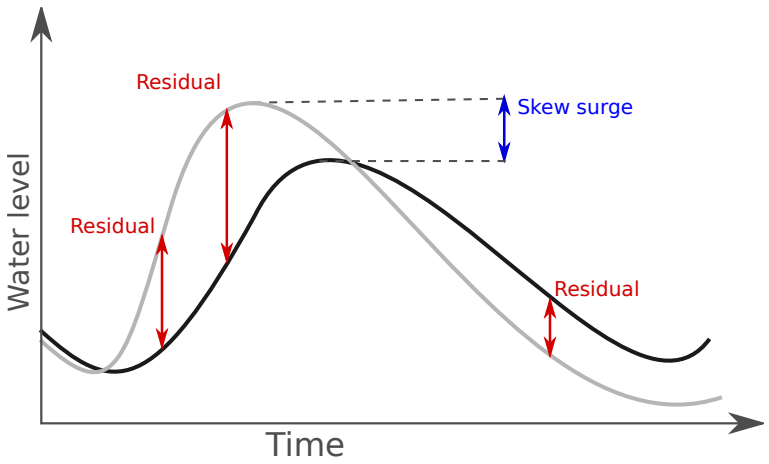

Figure 1: Schematic illustrating the difference between residual and skew surge. The black line shows the expected tide-only water level, and the grey line shows the observed/modelled water level including the atmospheric surge component (i.e. 'Total water level'). Residual is the difference between the two at any given time (red labels), and skew surge is the difference in peak water level regardless of its timing (blue label).

© Crown Copyright 2017

3.1 Observations

In order to compare the forecast performance of NEMO-Surge against that of CS3X, it is necessary to compare both models with observations. The observations used are tide gauge data from the UK National Class A Tide Gauge Network (National Tidal and Sea Level Facility, 2015), at the locations shown in Figure 2. Data are received in near-real time, with a temporal resolution of 15 minutes.

Figure 2: Class A tide gauge locations

These observations are received without any quality checks or processing having been done. Inspection of the data shows that there are occasions with clearly incorrect values—examples are shown in Figure 3. The raw observations have therefore been filtered before being used for model validation. This was done in two stages:

- An automated filter to remove any observations that are more than 3 standard deviations away from the mean, as calculated over a centered running window of 24 hours.
- Manual flagging of suspicious points during the analysis period that were not picked up by the previous process

In addition, for Wick only, all points for December 2016–January 2017 inclusive were flagged due to

© Crown Copyright 2017

the tide prediction file not being available for these dates.

The flagged points are masked out and are not used in the further analysis. The number of points that are excluded are shown in Table 1 (for those ports not included in the table, there were no excluded points). The excluded points comprise less than 1% of the 17472 total data points at all ports except Ullapool (1.2%), Immingham (3.6%), and Wick (34.1%).

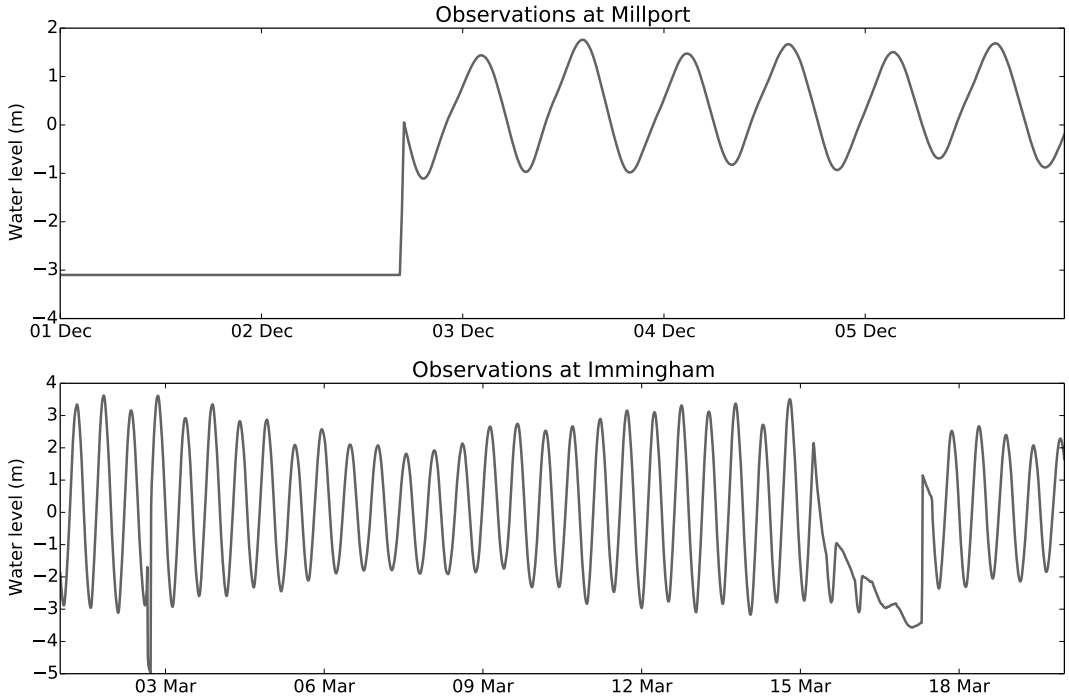

Figure 3: Example of unfiltered observations at Millport and Immingham showing some problem datapoints

The tide gauge observations are returned as total water level. These are compared with the synthesised total water level from the models (i.e. the model residual + harmonic tide prediction). Similarly, observed residual is calculated by subtracting the harmonic tide prediction from the tide gauge observation and then compared with the model residual.

It is known (Flather et al., 2000) that this introduces biases due to an element of 'double counting' of some long-period components. Briefly, the model tidal forcing does not include components longer than around a day, so longer constituents (such as *Sa* solar annual and *Ssa* solar semi-annual) are regarded as part of the surge and are included within the model residual. But these components are included in the harmonic tide prediction, so when this is added back on to the model residual they are effectively included twice within the model total water level.

Similarly, when the tide prediction is removed from the observed total water level to calculate the observed residual, these components are removed, although they remain in the model residual that the observation is compared with. This introduces biases into the comparisons.

Additionally, a recent project (Byrne et al., 2017) looking at surge prediction at Newport found a

© Crown Copyright 2017

Port	#
Aberdeen	2
Fishguard	100
Ilfracombe	3
Immingham	632
Kinlochbervie	5
Lerwick	1
Milford Haven	93
Millport	161
Mumbles	37
Newlyn	13
Plymouth	2
Ullapool	218
Weymouth	4
Whitby	2
Wick	5963
Workington	3

Table 1: The number of observations excluded from analysis for each port. For ports not listed, the number is 0

significant trend (order 1 cm/year) in the tide gauge observations, which is believed to be too large to be explained by sea level rise alone (Figure 4). The cause is currently unknown, but is possibly due to issues such as siltation of the instrument. This trend in sea-level introduces a further bias into the comparison as the tide harmonic prediction is calculated using a decades-long timeseries of observations. It is essentially then referenced to the average sea level during that time, which is not representative of the present observations. Byrne et al. (2017) found that the size of this offset was 15 cm at Newport—large enough to potentially account for all of the model-observation bias there.

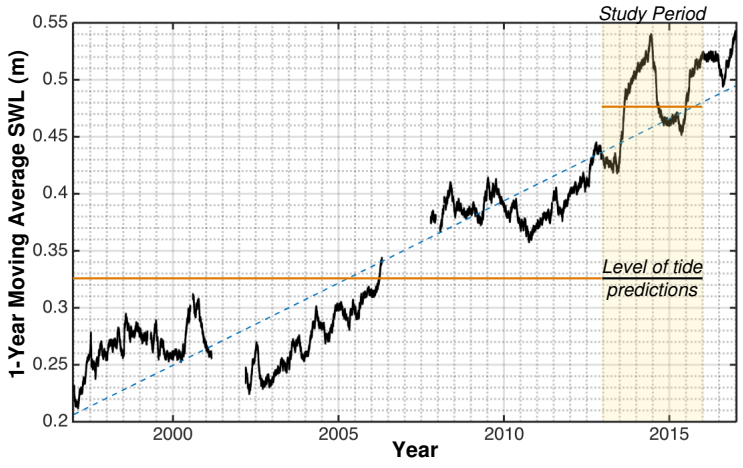

Figure 4: 1-Year moving average of Newport tide gauge observations between 1993–2016 (solid black line) and linear fit to all the data (blue dashed line). The lower orange horizontal line shows the calculated 'Z0' value that is the baseline used by the tide harmonic analysis [Reproduced with permission from Byrne et al. (2017)]

Although these biases could be corrected for, we have not done so in this analysis in order to match the outputs that are received by customers and used by them (noting that they do their own in-house

© Crown Copyright 2017

corrections). In any case these biases are identical for both CS3X and NEMO-Surge, and therefore do not affect the comparison between the two models. The focus of the project in its current stage is to match CS3X performance, to enable a smooth transition to the new model outputs for customers. In future updates, when we aim to introduce improvements to the model, it will be necessary to account for these biases to avoid fitting the model to the wrong solution.

3.2 Qualitative analysis (case studies)

For the case studies, results from individual model forecast runs are presented in two formats:

- Timeseries forecast from a single model cycle, showing the ensemble distribution and how this compared with observations
- Box-and-whisker plots for a number of model cycles in the lead up to the case study event showing the ensemble spread of maximum residual and skew surge. For each model cycle, the peak residual and skew surge that occur close to the event time are identified using a search window of 11 hours centred on the observed event time, then a box-and-whisker plot is added to show the ensemble spread. This analysis shows the evolution of the predicted magnitude of the surge over subsequent forecasts regardless of any changes in the exact timing between different forecasts, or between ensemble members. This process is illustrated by Figure 5

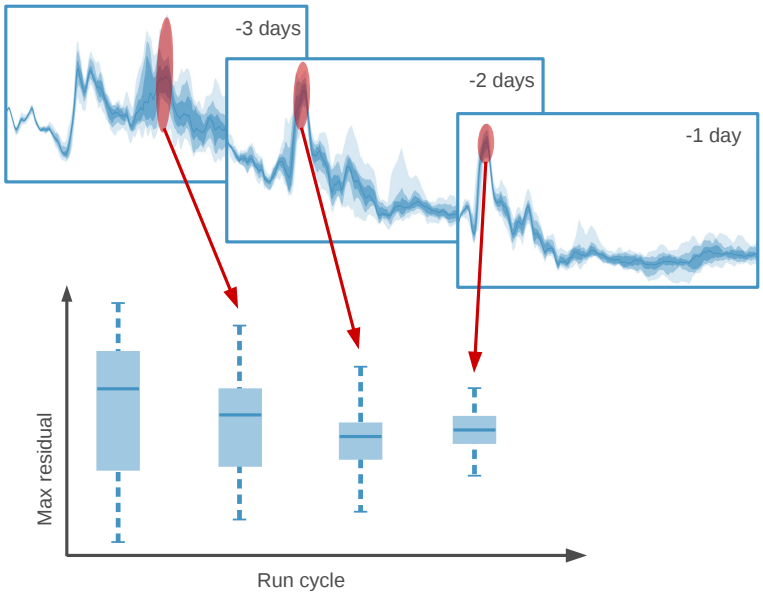

Figure 5: Schematic illustrating how a timeseries of box plots is created from a series of forecasts. For each individual forecast, the peak residual is found for each ensemble member, then a box-and-whisker plot is plotted for that range of values. The same procedure is also applied for skew surge.

All ports were reviewed when examining the case studies, but in this report only key examples with particularly high residuals or peak water levels are shown. As a result of the type of events which

© Crown Copyright 2017

occurred during this period and the location of areas with particularly shallow waters around the UK, the majority of the examples are at ports of primary interest to the EA.

3.2.1 Note on storm names

The 2016–17 winter was the second year in the Met Office and Met Éirann joint pilot project to provide official names for the most severe winter storms, with the aim of engaging the media and public and increasing reach of severe weather warnings. In the first season names were only issued on the basis of wind impacts, but this was extended in 2016–17 to also include rain and snow. A storm is allocated a name if there is deemed to a be a risk of 'medium' or 'high' impacts, as defined by either or both of the Met Office's and Met Éirann's severe weather impact tables.

2016–17 was a relatively quiet season with only 5 storms being issued names compared with 11 in the 2015–16 season, despite the extension in naming criteria (UK Storm Centre, 2017). Not all the cases examined in this report are named storms.

3.3 Quantitative analysis (overall statistics)

Model performance statistics have been calculated over the period 1 December 2016–31 May 2017. In order to combine the various model runs over this period, and to allow comparison at different lead times, the forecasts have been binned into 6-hour long leadtime 'slots'. This bin size is chosen as the model runs are 6 hours apart, so this produces a gap-free timeseries for each leadtime slot.

The binning process is illustrated by Figure 6: for each model run (shown by the different colours), the first 6 hours are placed into the first leadtime bin, the second 6 hours are put into the second leadtime bin, and so on (i.e. moving along the horizontal axis in the figure). Then for the next model cycle the same thing is done, but now the valid time represented by each leadtime bin is 6 hours later than the previous run (i.e everything has moved up on the vertical axis in the figure). In this way for each leadtime we build up a synthetic timeseries composed of different model run cycles. This process is repeated for each ensemble member, and the ensemble mean is then computed.

Statistics are calculated for each tide gauge location by comparing the ensemble mean with the observations for each of the leadtime timeseries. An average is then also taken over all of the locations to provide an overall measure of model performance. The statistics used are:

- Mean error (bias)
- RMS error
- Ratio of observed and modelled standard deviation
- r^2

© Crown Copyright 2017

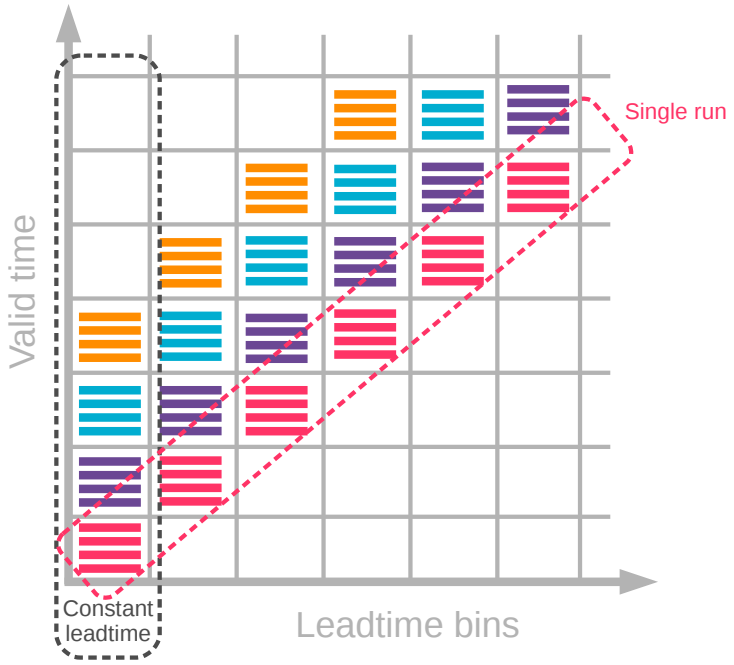

Figure 6: Schematic illustrating how the model forecasts are binned into leadtime slots. Different model run cycles are shown by different colours.

In this study the ensemble mean is focused on in order to asses any systematic differences between the surge models. Other forms of probabilistic forecast verification are ignored as they will mainly be influenced by atmospheric variations in the ensemble, and the forcing data are applied identically to the two surge models.

4 Case studies: North Sea coast

4.1 26 & 27 December 2016 (storm Conor)

Storm Conor brought 60–70 kt winds to northwest Scotland on 25–26 December, generating significant surge residuals in the early hours of both the 26 and 27 December. Figure 7 shows example wind and pressure fields from the control member of the MOGREPS-G forecast, for 00Z and 12Z on the 26th. These show strong westerly winds, shifting to north-westerly as storm Conor moves eastward across the northern part of the domain.

The storm centre travelled across the north of the UK and into Scandinavia, with the strong winds generating a surge in the north of the domain. The surge then propagated down into the North Sea, impacting the east coast of the UK, albeit with only a limited amount of forcing being applied to the surge within the North Sea basin. The path of the storm and the footprint of sites affected matches the ‘category 4’ footprint identified by Haigh et al. (2016) in their analysis of UK surge events over the last 100 years.

© Crown Copyright 2017

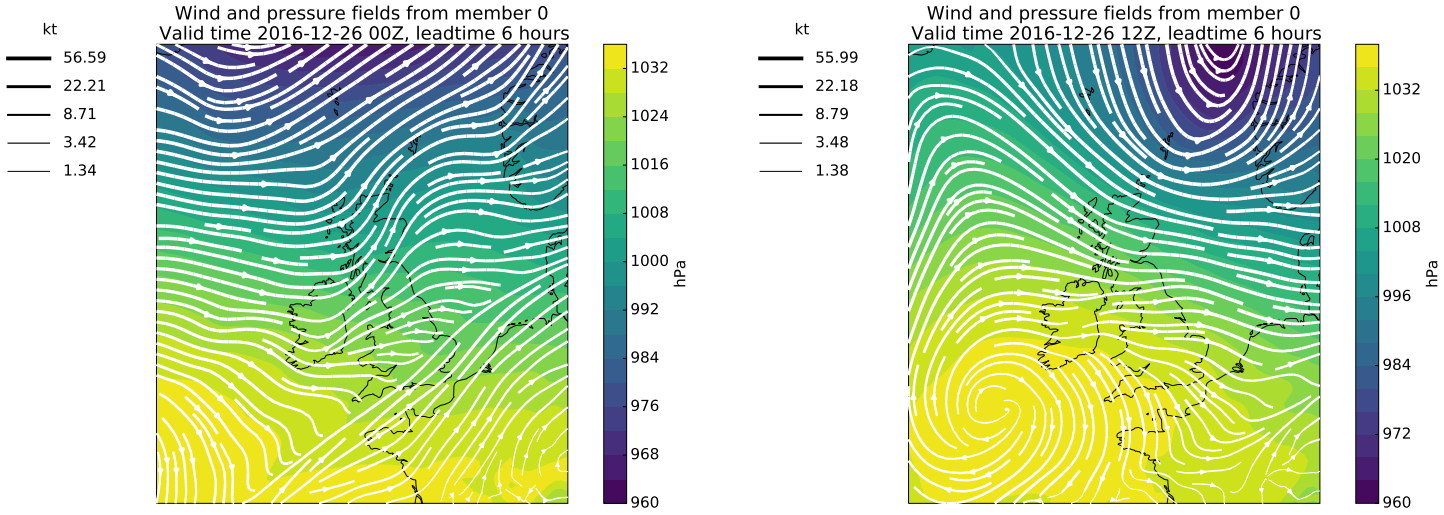

Figure 7: MOGREPS-G control member forecast wind (lines) and pressure fields (contours) for 00Z (left) and 12Z (right) on 26 December, at 6 hours lead time

As this surge was forced remotely rather than locally at the sites affected, we might expect there to be more difference between NEMO-Surge and CS3X as the propagation is more heavily influenced by model specific features such as the underlying bathymetry.

Although the residual surge was largest on the 27th, at most locations the total water levels were highest on the 26th as the residual peaked nearer to high tide. These events occurred on the run up to the spring tide with the new moon on 29 December.

Example surge forecasts for Cromer are shown in Figures 8 (NEMO) and 9 (CS3X). We see that the models are generally similar, with both generally over-forecasting the residual on both days. CS3X is rather worse on the first residual peak than NEMO-Surge at this location. NEMO-Surge then has a better drop off rate for the residual, leading to a slightly better forecast at high tide.

The box plots in Figures 10–13 show that both models provided several days' warning of these surges. In particular, for the first surge on the 26th there was clear indication from the models more than a week in advance, albeit with a larger range of uncertainty over the magnitude until around 3 days before. For the second event on the 27th there was at least 3–4 days warning of the surge.

NEMO-Surge and CS3X both show very similar results for the residual and skew surge predictions over the range of lead times. For example, observed skew surge falls within the lower percentiles of the ensemble for both models at lead times of 5 days ahead, or less, for events on both 26th and 27th December. Both models under-estimate the maximum residual on the 27th at lead times of 6–7 days ahead.

© Crown Copyright 2017

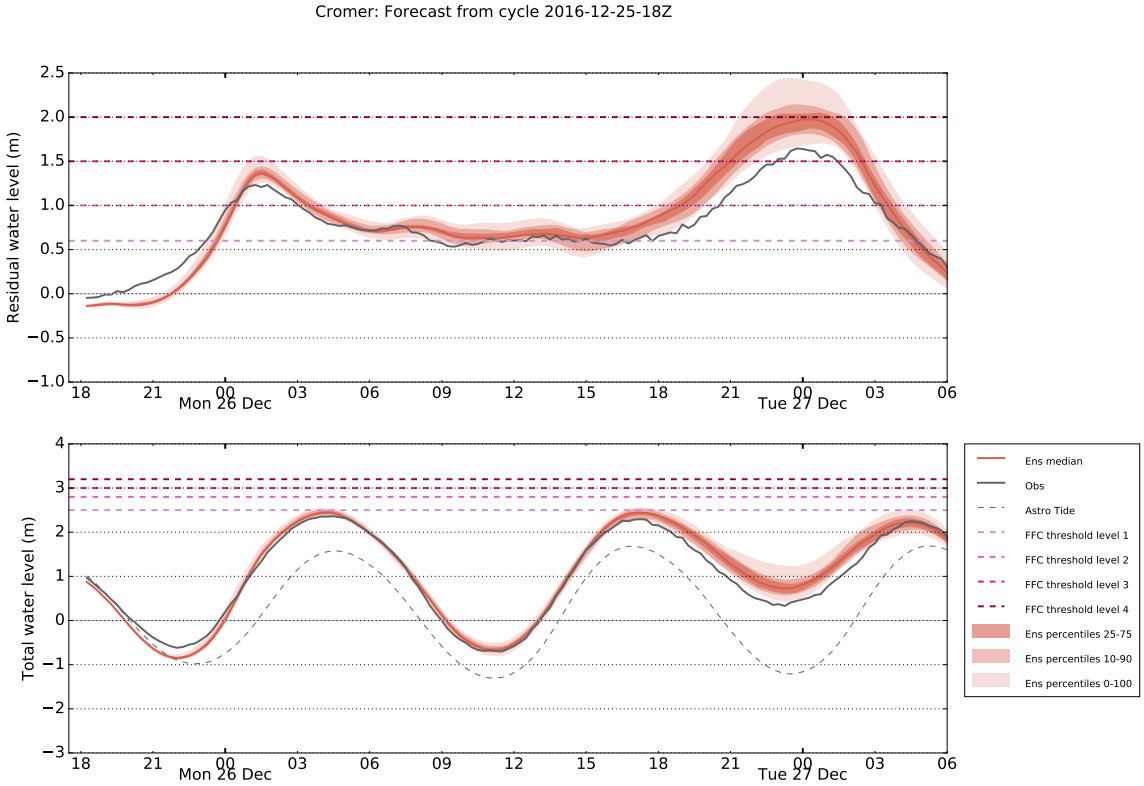

Figure 8: NEMO-Surge forecast for Cromer from 2016-12-25 18Z run. Coloured areas show the forecast ensemble, the solid grey line shows observations, and the dashed grey line shows the tide prediction.

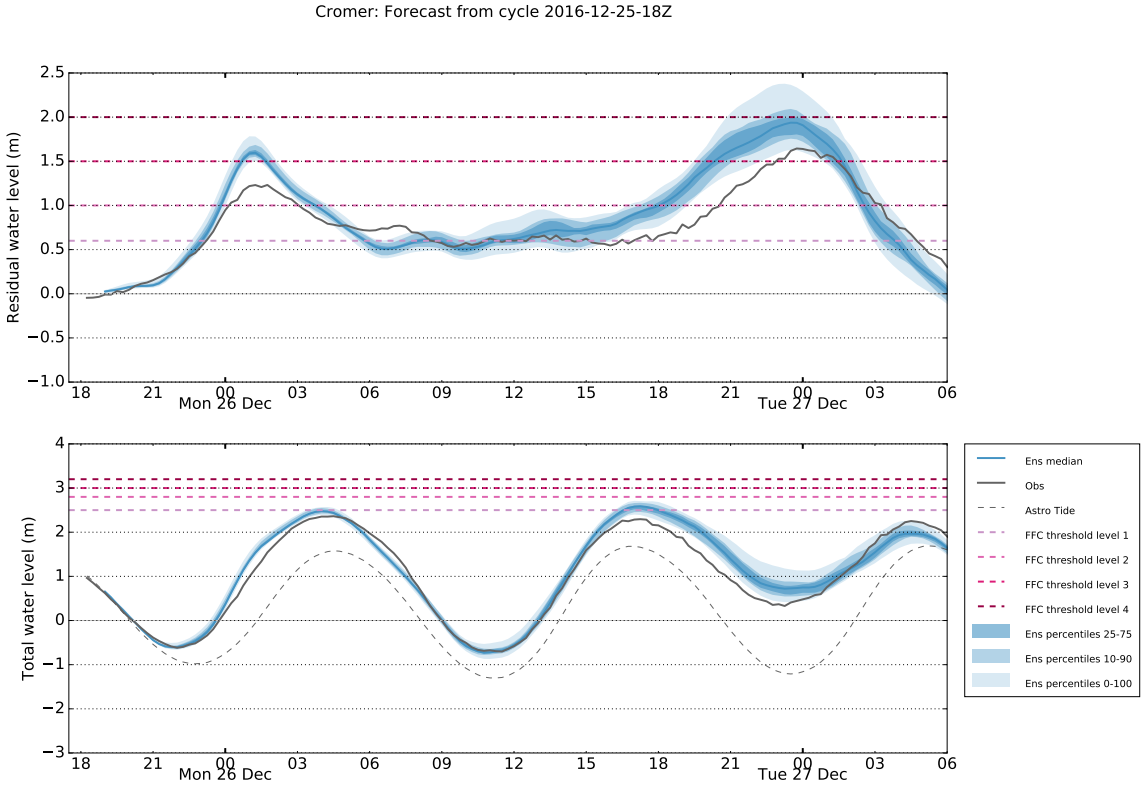

Figure 9: CS3X forecast for Cromer from 2016-12-25 18Z run

© Crown Copyright 2017

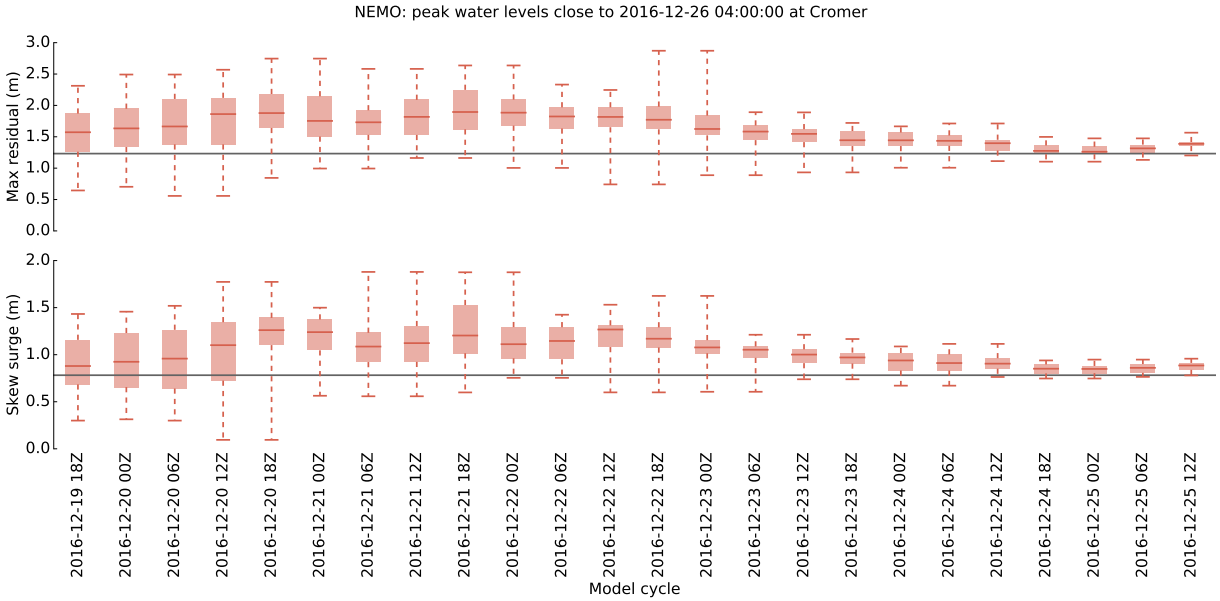

Figure 10: Boxplots showing NEMO-Surge ensemble distribution of peak residual and skew surge for the surge on 26 December. The coloured box and whiskers show the median, 25–75 percentiles, and 0–100 percentiles from the model ensemble. The grey line shows the observed value.

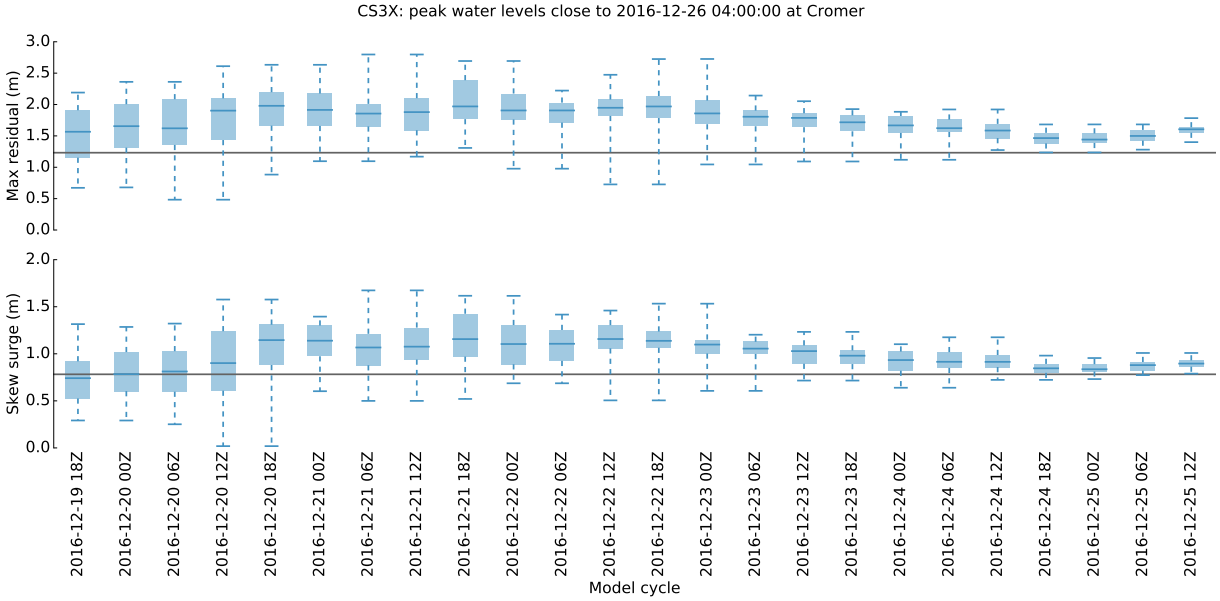

Figure 11: Boxplots showing CS3X ensemble distribution of peak residual and skew surge for the surge on 26 December. The coloured box and whiskers show the median, 25–75 percentiles, and 0–100 percentiles from the model ensemble. The grey line shows the observed value.

© Crown Copyright 2017

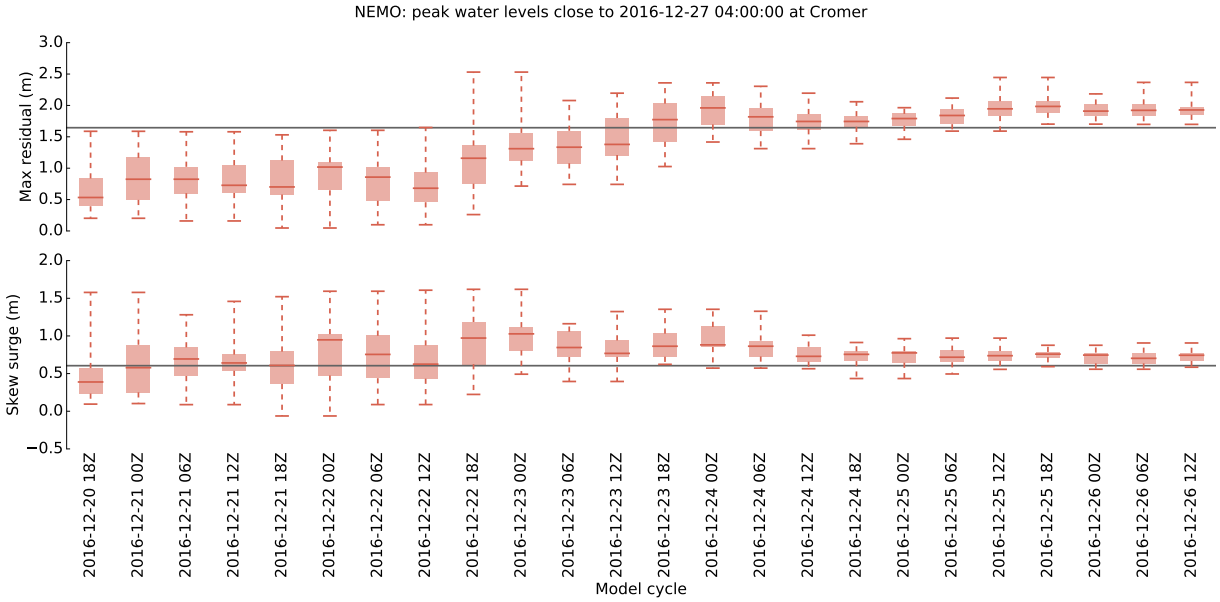

Figure 12: Boxplots showing NEMO-Surge ensemble distribution of peak residual and skew surge for the surge on 27 December. The coloured box and whiskers show the median, 25–75 percentiles, and 0–100 percentiles from the model ensemble. The grey line shows the observed value.

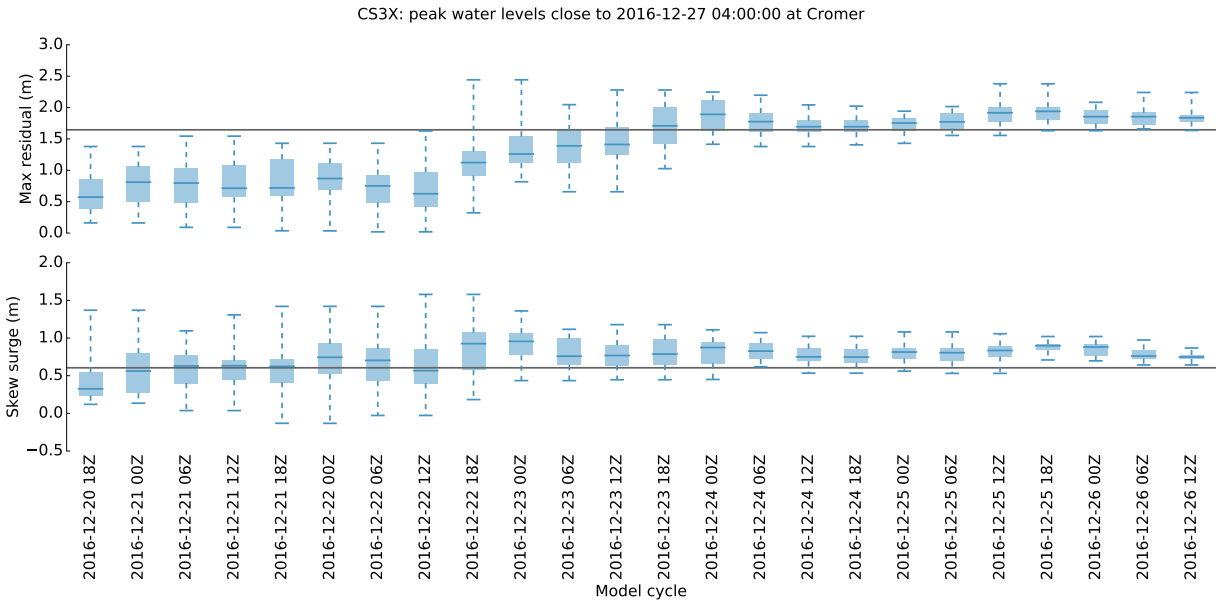

Figure 13: Boxplots showing CS3X ensemble distribution of peak residual and skew surge for the surge on 27 December. The coloured box and whiskers show the median, 25–75 percentiles, and 0–100 percentiles from the model ensemble. The grey line shows the observed value.

© Crown Copyright 2017

4.2 13 January 2017

This storm coincided with spring tides, coming the day after the full moon on 12 January. Although the storm itself did not meet the threshold criteria to be given a name, the resulting surge was the most significant during the season in terms of impacts, including flooding at Whitby during the early evening. There was also particular concern for the Essex coast and residents in the town of Jaywick (where more than 30 people were killed in the 1953 storm surge) were urged to evacuate their homes ahead of the evening high tide.

Figure 14 shows example MOGREPS-G forecast fields for 00Z and 12Z on the 13th showing sustained north-westerly to northerly winds blowing down the North Sea. This is in contrast with storm Conor previously discussed, as this time the surge was forced more locally within the North Sea. As both models use the same atmospheric forcing, we would therefore expect the model surge forecasts to be more similar than is the case for the storm Conor surge.

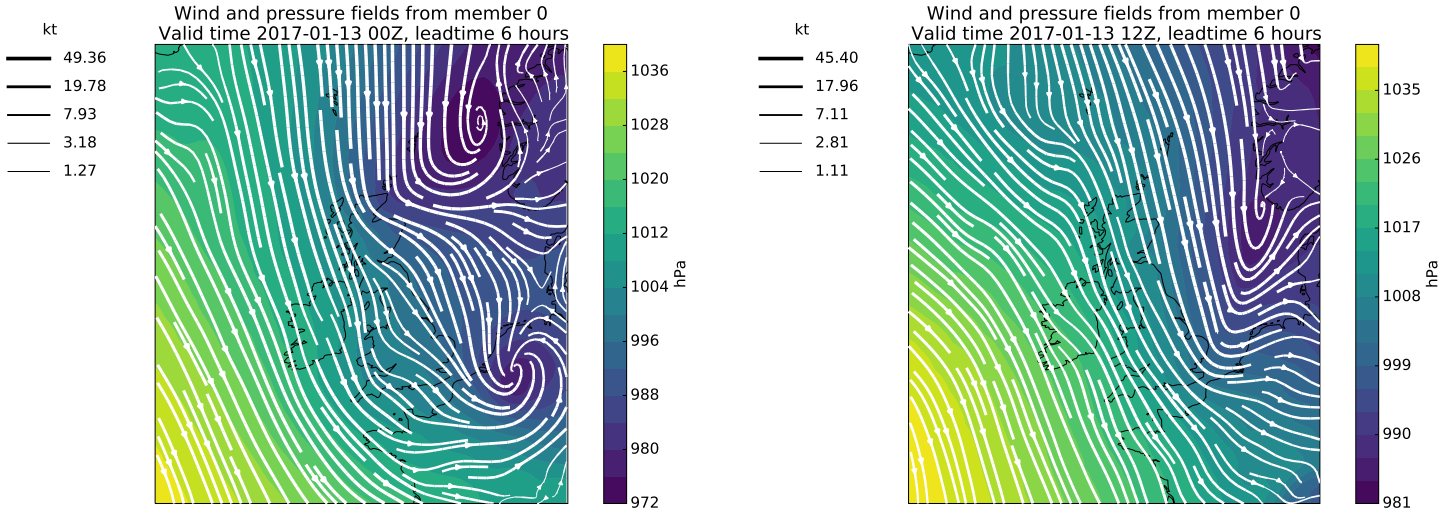

Figure 14: MOGREPS-G control member forecast wind (lines) and pressure fields (contours) for 00Z (left) and 12Z (right) on 13 January, at 6 hours lead time

Looking first at Whitby, the boxplots in Figures 15 and 16 show that the ensemble indicated the potential for large surge values around 4 days in advance of the event, with the ensemble spread narrowing in as the leadtime shortened. As expected the models both show very similar patterns. NEMO-Surge was slightly closer to the observed levels but the difference is small.

Looking at individual run cycles, the earlier forecasts had a spread of large residuals impacting both morning and evening high tides (eg Figure 17). This narrowed down in forecasts issued nearer the time to predict that the largest impact would be on the evening high tide (eg Figures 18,19).

The 00Z forecast on the 13 January (which would have been available to forecasters around 7am) from both NEMO-Surge and CS3X predicted the residual surge to peak at around 11am, coinciding with low tide. The model residual then dropped off as high tide approached, but still remained high enough for the total water level to breach the top threshold level.

© Crown Copyright 2017

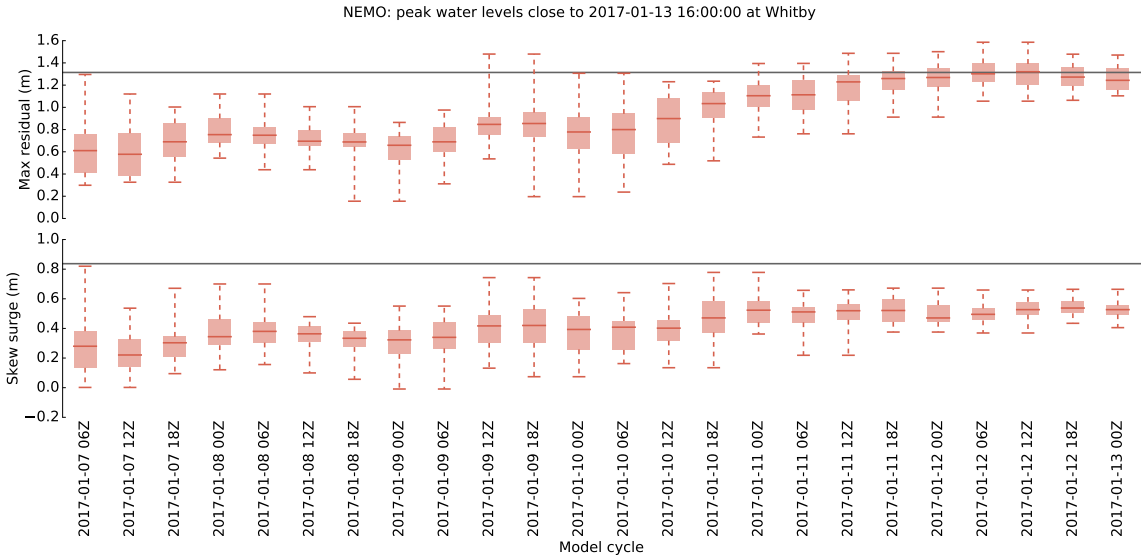

Figure 15: Boxplots showing NEMO-Surge ensemble distribution of peak residual and skew surge at Whitby for the surge on 13 January. The coloured box and whiskers show the median, 25–75 percentiles, and 0–100 percentiles from the model ensemble. The grey line shows the observed value.

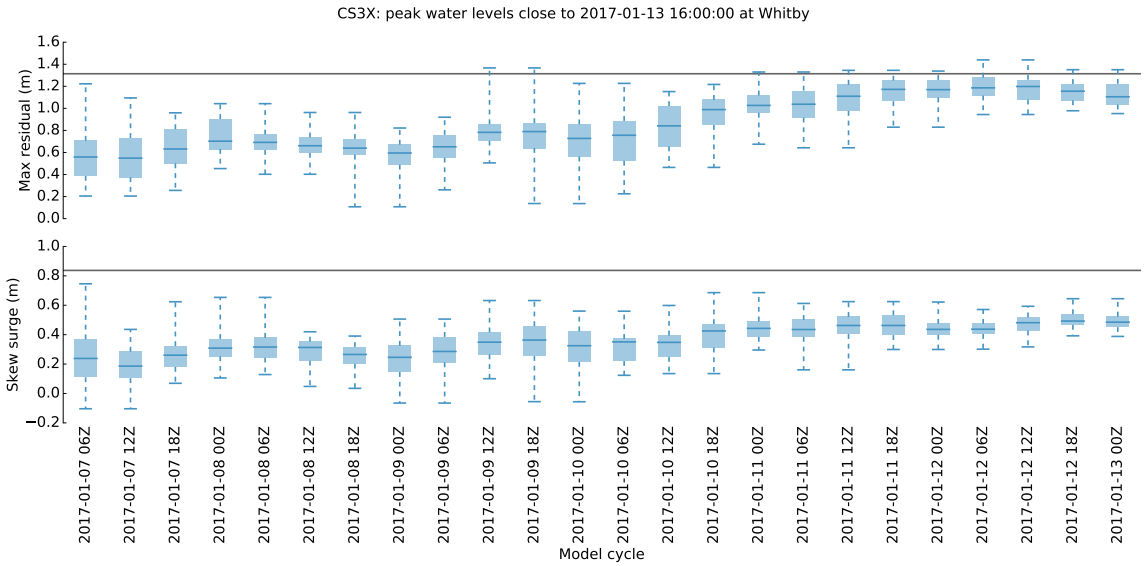

Figure 16: Boxplots showing CS3X ensemble distribution of peak residual and skew surge at Whitby for the surge on 13 January. The coloured box and whiskers show the median, 25–75 percentiles, and 0–100 percentiles from the model ensemble. The grey line shows the observed value.

© Crown Copyright 2017

Comparing with the observed water level (shown by the solid grey line in the figures), the peak residual was reasonably well forecast with the observed level falling within the ensemble range for both models. For NEMO-Surge the observed peak was slightly closer to the central distribution than it was for CS3X. In both models the residual then dropped off too quickly, leading to an under-predicted peak water level at the evening high tide. Forecast uncertainly remained quite large over the forecast with quite a lot of spread in the ensemble.

Whitby is a site that is known to have a model bias compared with observations and forecasters systematically bias correct the forecasts (Flood Forecasting Centre, personal communication).

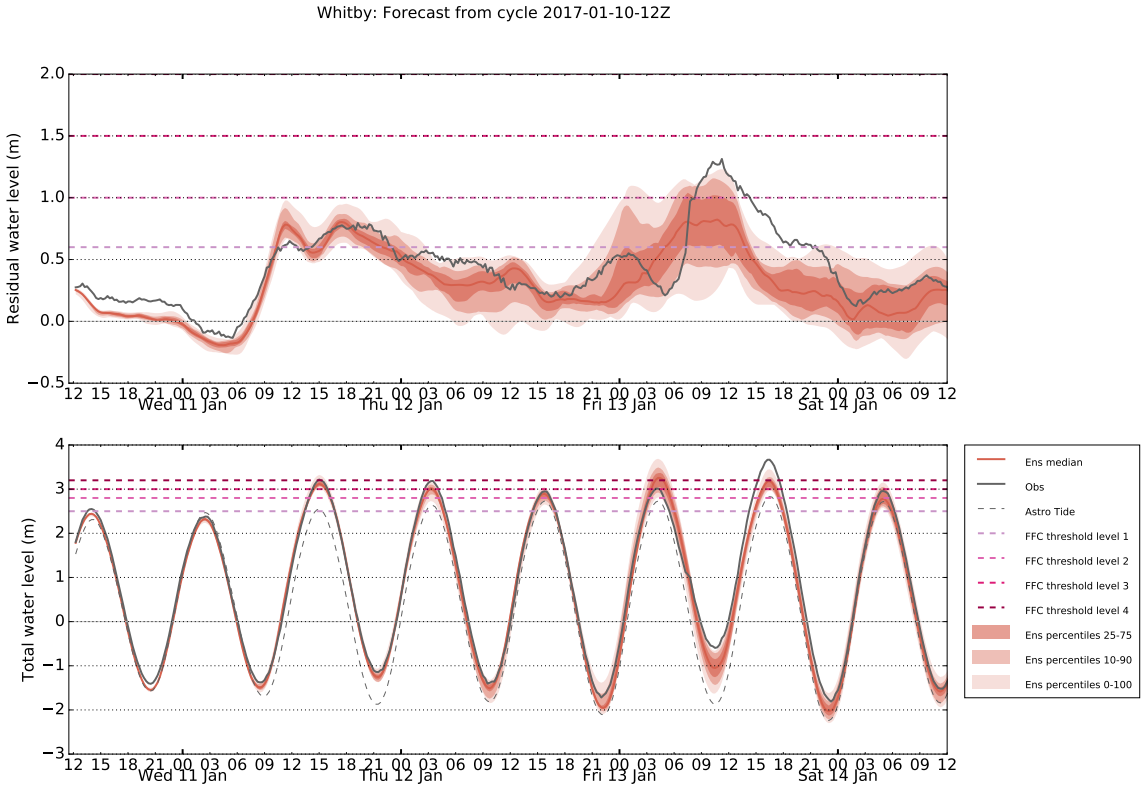

Figure 17: NEMO-Surge forecast for Whitby from 2017-01-10 12Z run, showing forecasts at this leadtime indicated potential for a large surge coinciding with both of Friday's high tides

© Crown Copyright 2017

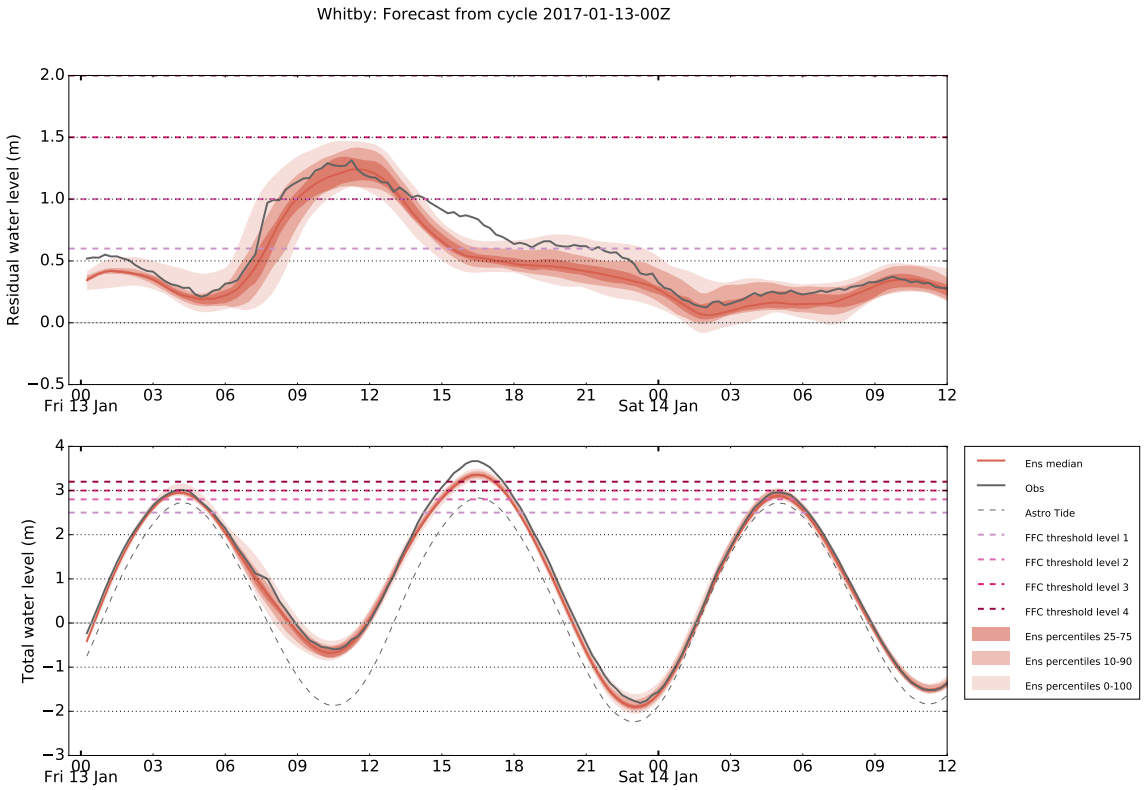

Figure 18: NEMO-Surge forecast for Whitby from 2017-01-13 00Z run. At this shorter leadtime, focus has shifted to the evening high tide with less concern for the morning.

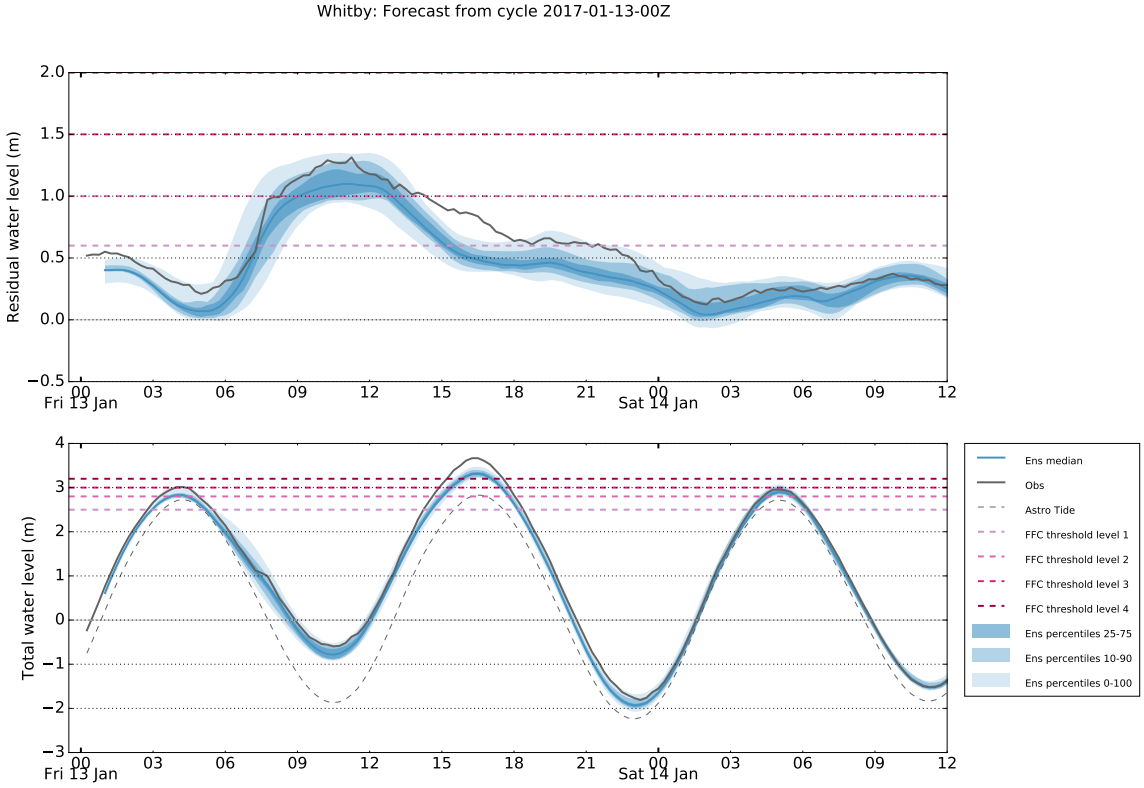

Figure 19: CS3X forecast for Whitby from 2017-01-13 00Z run

© Crown Copyright 2017

The pattern in East Anglia and Essex was rather different with both models *over* estimating the maximum residual. At Cromer and Sheerness, both models came back in line with observations around high tide, so that the maximum total water level observed was within the forecast ensemble. For example, Figures 20 and 21 show the forecasts for Sheerness.

However, at Harwich (which is the closest tide gauge location to Jaywick), shown in Figures 22 and 23, the residual continued to be over-forecast, leading to the total water level also being too high. This was a difficult situation for forecasters as the observed water heights were tracking close to (or even above) the model ensemble median for much of the day, when decisions were being made, before stalling ahead of the evening high tide. Fortunately for residents, the residual had dropped sufficiently by high tide to avoid significant flooding impacts in this region.

For both Sheerness and Harwich the model timings of peak water level show some small differences, reflecting effects of bathymetry and tidal propagation in the two surge models. Based on results from Furner et al. (2016), we might expect the NEMO-Surge solution for Harwich to be better and, conversely, CS3X to be better at Sheerness. Qualitatively, the observed water levels in this case agree with this expectation. Nevertheless, the overall skew surge for each high tide is very similar.

© Crown Copyright 2017

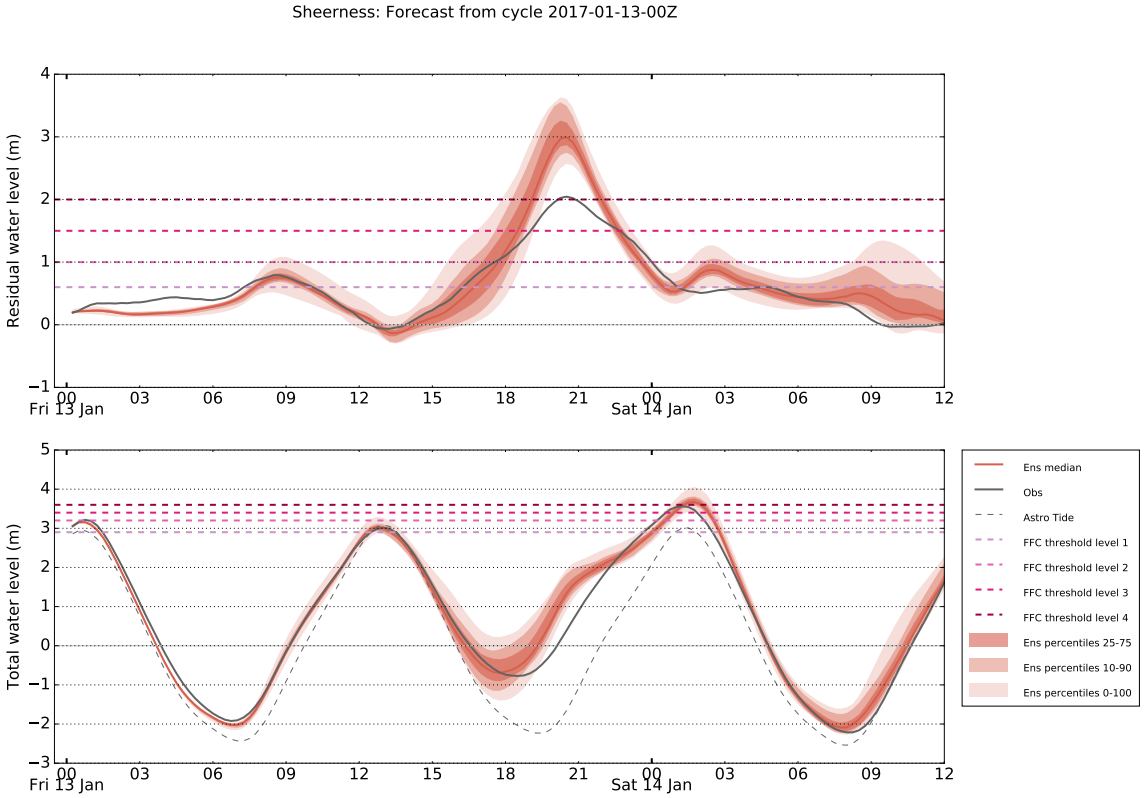

Figure 20: NEMO-Surge forecast for Sheerness from 2017-01-13 00Z run

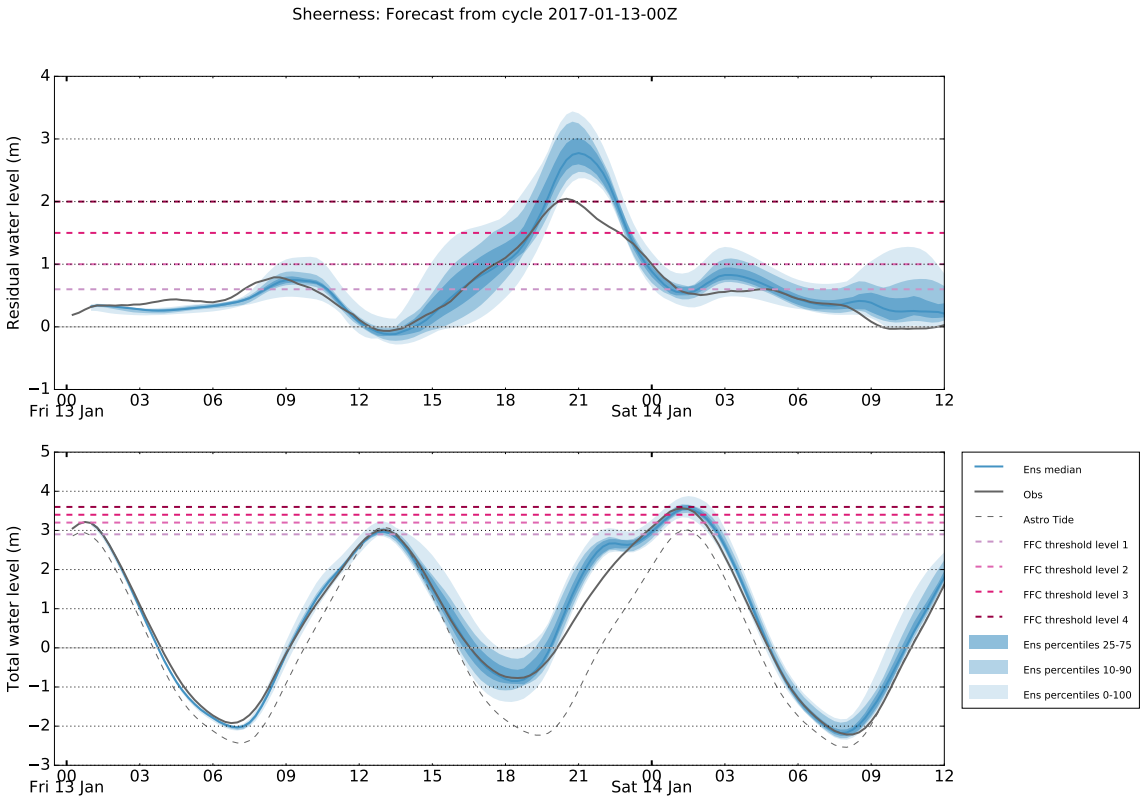

Figure 21: CS3X forecast for Sheerness from 2017-01-13 00Z run

© Crown Copyright 2017

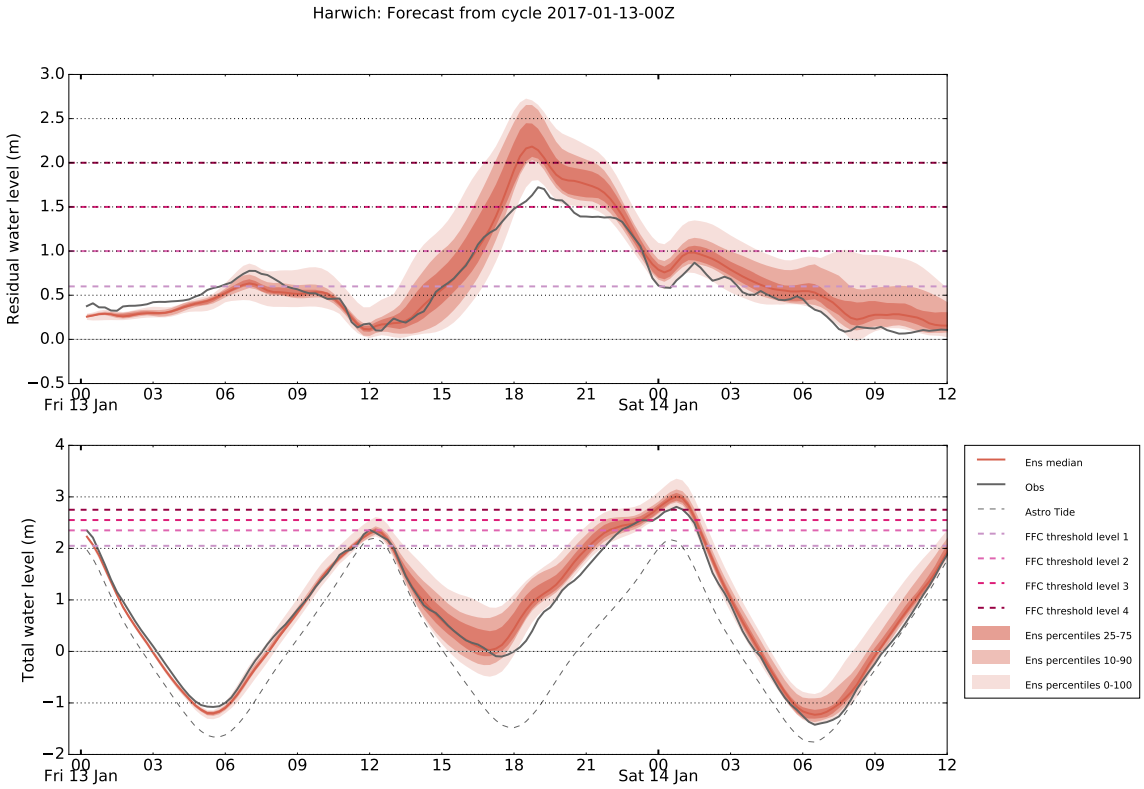

Figure 22: NEMO-Surge forecast for Harwich from 2017-01-13 00Z run

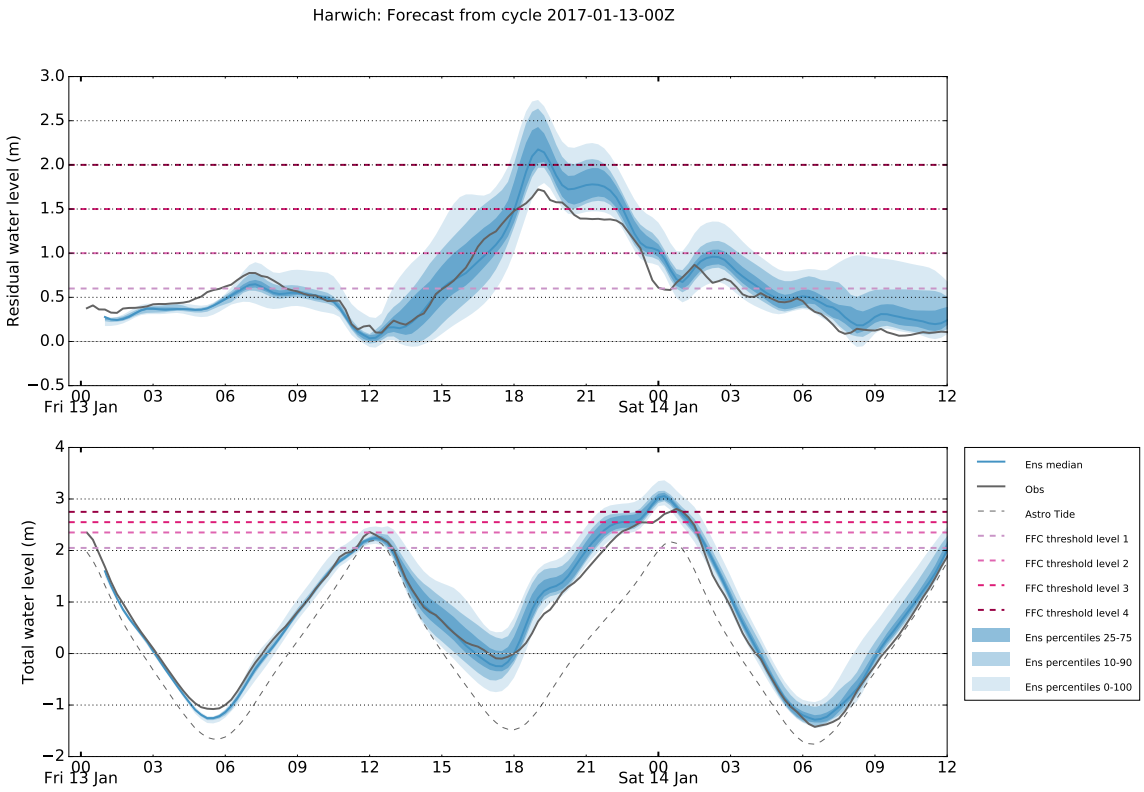

Figure 23: CS3X forecast for Harwich from 20170113 00Z run

© Crown Copyright 2017

5 Case studies: Irish Sea coast

5.1 23 December 2016 (storm Barbara)

Between 23–24 December, storm Barbara brought strong winds to much of the UK, especially north and northwest Scotland where gusts of 60–70 kt were recorded at low levels. This came a few days after neap tides on the 21st so, although the residual peaked close to high tide in most locations, the total water level was still well below threshold levels and no impacts were recorded. Figure 24 shows the forecast wind field at 00 and 12Z on the 23rd, with southwesterly winds blowing directly up the Irish Sea. The largest surge residuals were seen around the northern half of the Irish Sea at Llandudno, Liverpool, Heysham, and Workington. This spatial pattern and the storm track fit well within Haigh et al. (2016)'s 'category 2' surge type.

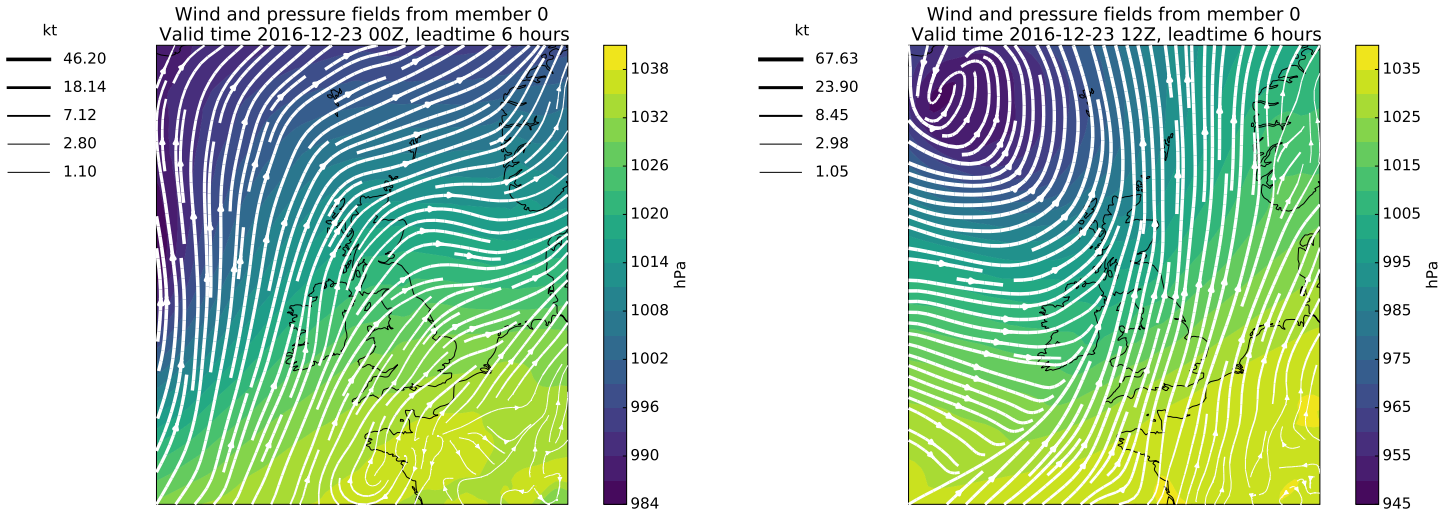

Figure 24: MOGREPS-G control member forecast wind (lines) and pressure fields (contours) for 00Z (left) and 12Z (right) on 23 December (storm Barbara), at 6 hours lead time

Figures 25 and 26 show an example surge forecast for Workington. We see that both models provided a good forecast for the increasing residual and the maximum peak level, but they both remained too high after peaking compared with the observations which fell away more quickly. NEMO-Surge and CS3X are similar, except that NEMO-Surge seems to have a slightly better shape for the residual.

At Liverpool (Figures 27 and 28), NEMO-Surge slightly underestimates the maximum residual in this case compared with CS3X. Again this is likely due to phase differences in the model tide, the effects of which are minimised at high water so that there is very little difference in peak total water level and skew surge.

© Crown Copyright 2017

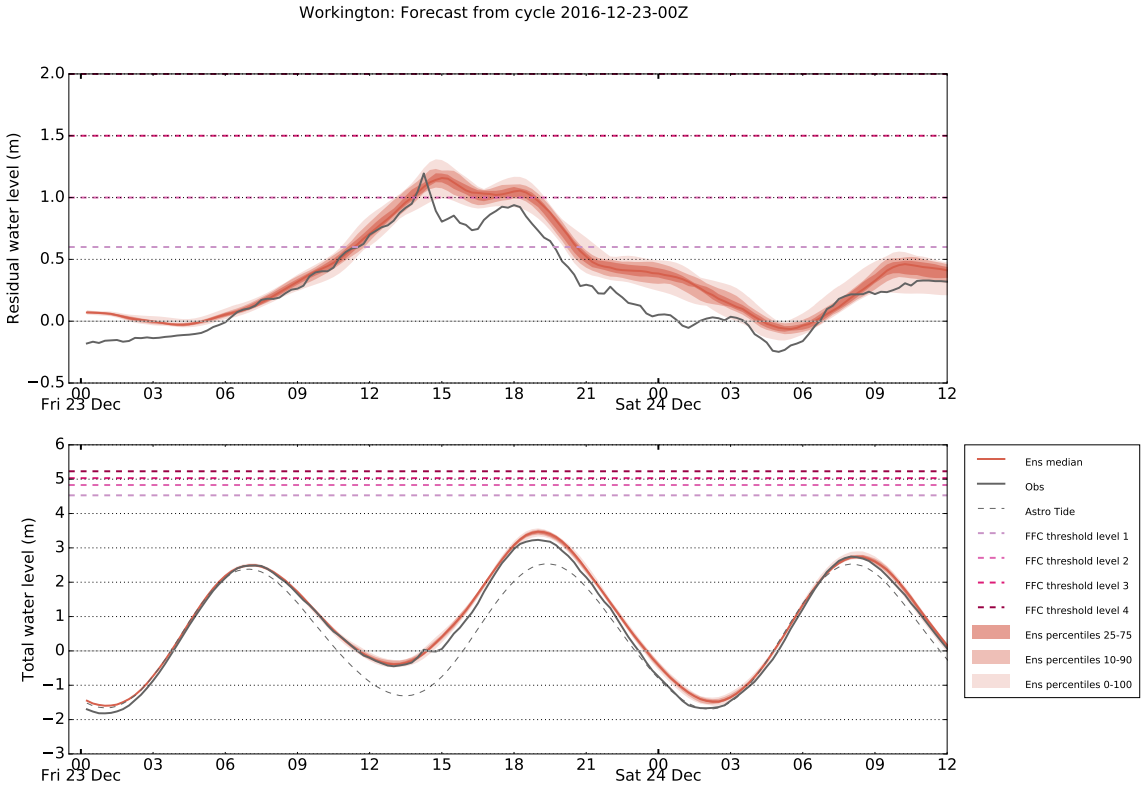

Figure 25: NEMO-Surge forecast for Workington from 2016-12-23 00Z run

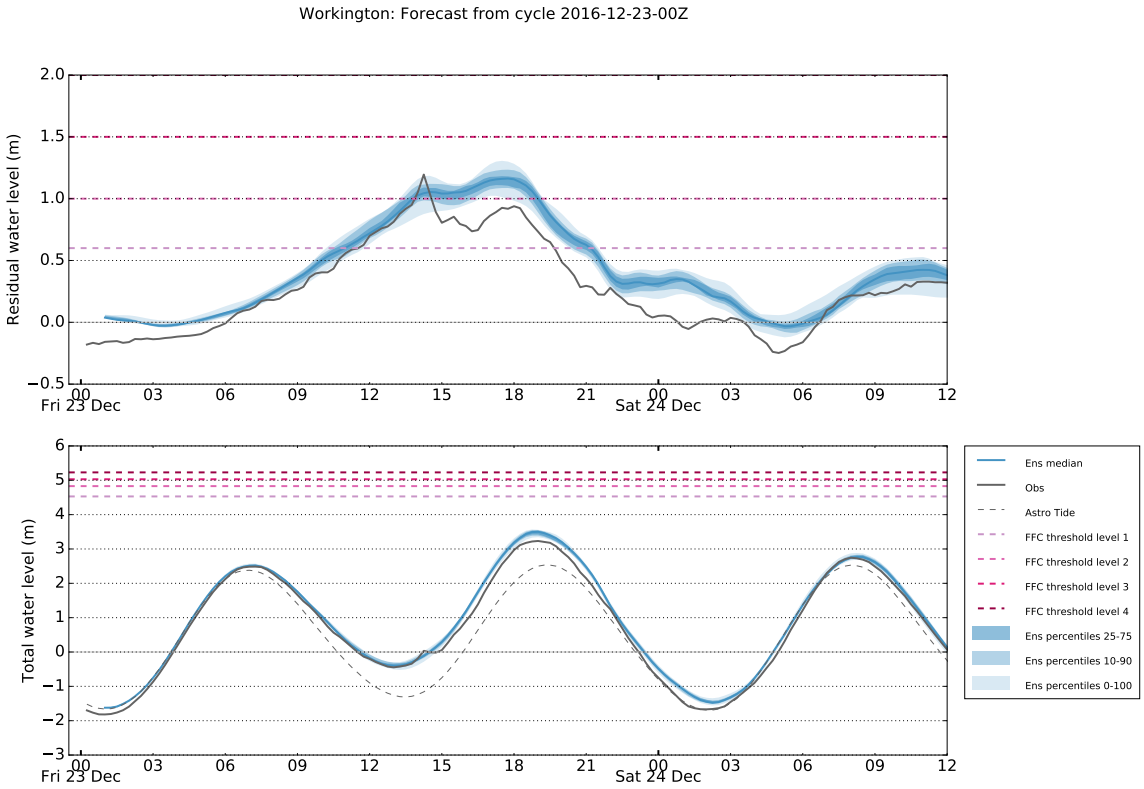

Figure 26: CS3X forecast for Workington from 2016-12-23 00Z run

© Crown Copyright 2017

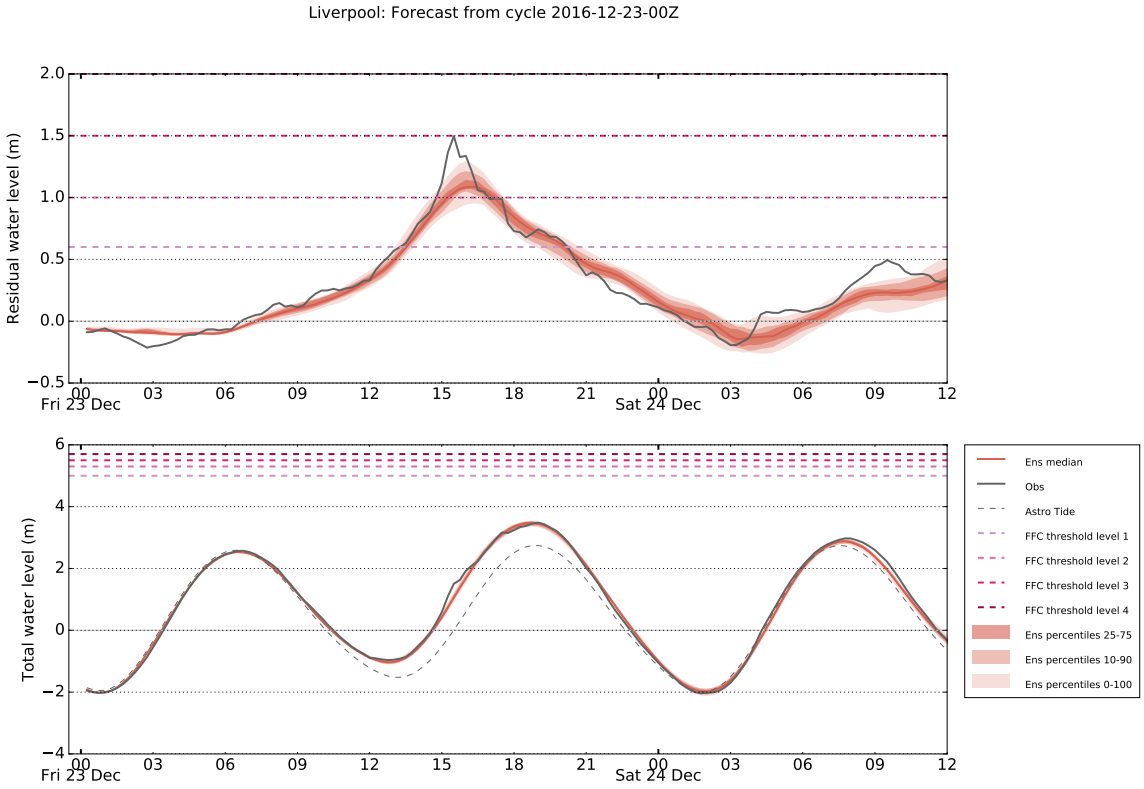

Figure 27: NEMO-Surge forecast for Liverpool from 2016-12-23 00Z run

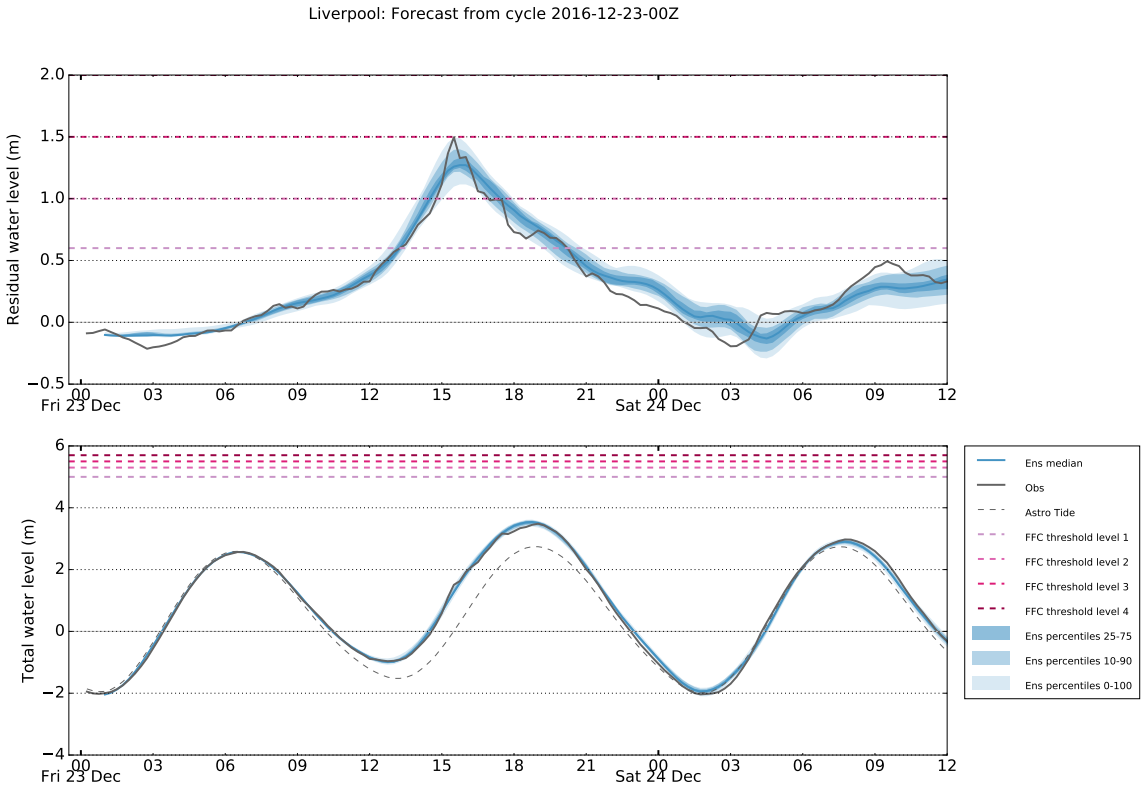

Figure 28: CS3X forecast for Liverpool from 2016-12-23 00Z run

© Crown Copyright 2017

5.2 23 & 26 February 2017 (storms Doris and Ewan)

Storm Doris hit the UK on 23 February with yellow and amber warnings issued for wind, snow, and rain for large parts of the UK. Peak wind gusts reached 82 kt in Capel Curig (Wales), with 50–60 kt being widespread. The impacts of the wind in particular were significant, including 3 deaths and widespread disruption to transport. The surge impacts were less significant, but did include reports of flooding at Cork. Shortly after Doris, on 26 February a second (less severe) storm named Ewan brought strong winds again to the south of Ireland.

Figures and 29 and 30 show example MOGREPS-G wind fields for Doris and Ewan respectively. The surge footprint from both storms is most similar to Haigh et al. (2016)'s 'category 2' with Irish Sea ports being predominantly affected. However the storm track taken by Doris is somewhat unusual for this category, going directly across the UK along a track that would usually tend to affect the South coast and Bristol Channel ports more than those in the Irish Sea ('category 1').

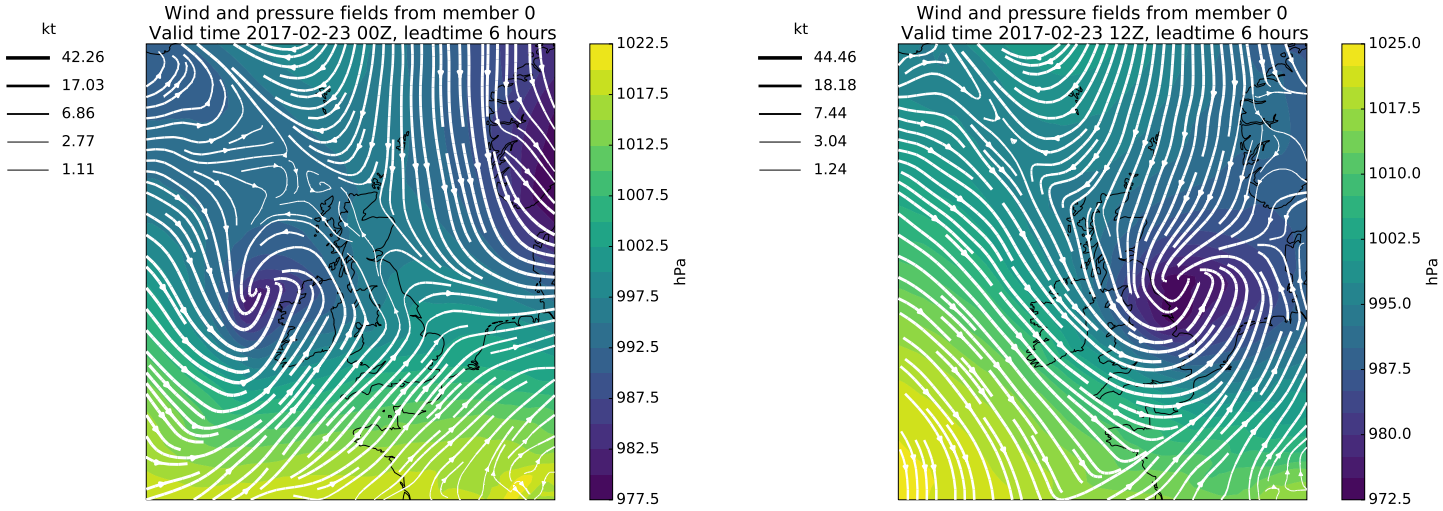

Figure 29: MOGREPS-G control member forecast wind (lines) and pressure fields (contours) for 00Z (left) and 12Z (right) on 23 February (storm Doris), at 6 hours lead time

Doris occurred in the middle of the spring-neap cycle so, although the residuals were generally largest quite close to high tide, the total water levels did not breach warning thresholds. Ewan coincided with spring tides (new moon 26 February), but fortunately the residual generally peaked around low tide so the impact was again limited.

The largest residual observed was at Liverpool during storm Doris, at almost 2 m. This was poorly forecast by both CS3X and NEMO-Surge, with the models significantly underestimating the size of the surge (Figures 31 and 32). Further analysis, outside the scope of this report, would be needed to determine if the under-prediction is due to a poor atmospheric forecast or issues with the surge model schemes. NEMO-Surge and CS3X were however very similar to each other.

For storm Ewan a few days later (Figures 33 and 34), the residual surge at Liverpool was smaller and was better forecast, coming in at the top range of the model ensembles. Again NEMO-Surge

© Crown Copyright 2017

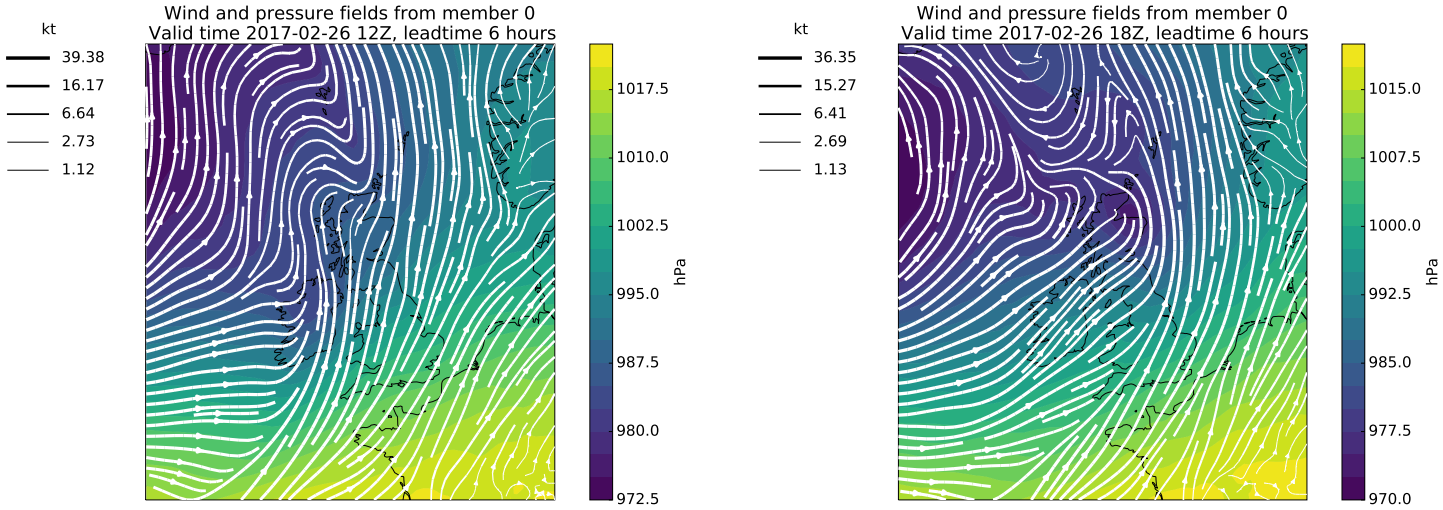

Figure 30: MOGREPS-G control member forecast wind (lines) and pressure fields (contours) for 12Z (left) and 18Z (right) on 26 February (storm Ewan), at 6 hours lead time

and CS3X are similar, although (in contrast to the previous case) NEMO-Surge has a slightly better timing and slope for the residual.

Figures 35 and 36 show the forecast for Barmouth 3 days in advance of storm Doris. Both models were very similar and showed a clear signal for elevated water levels, but the magnitude was again underestimated. As we approached the time of the storm however, the models both improved to have a much better residual peak (Figures 37 and 38). In this case NEMO-Surge still underestimated the residual for the first peak around 7am, leading to an under-estimate of peak water level. It then came in line with observations (and similar to CS3X) for the second, larger, peak around 10am.

Over the trial period as a whole, statistical analysis shows that there is little difference in performance between NEMO-Surge and CS3X at Barmouth. For example, Figure 39 shows the RMS error in both residual and skew surge for the model ensemble means calculated over the December–May period. We see that both models are extremely similar with maximum differences of only around 3 mm (with NEMO-Surge being the best at most lead times).

© Crown Copyright 2017

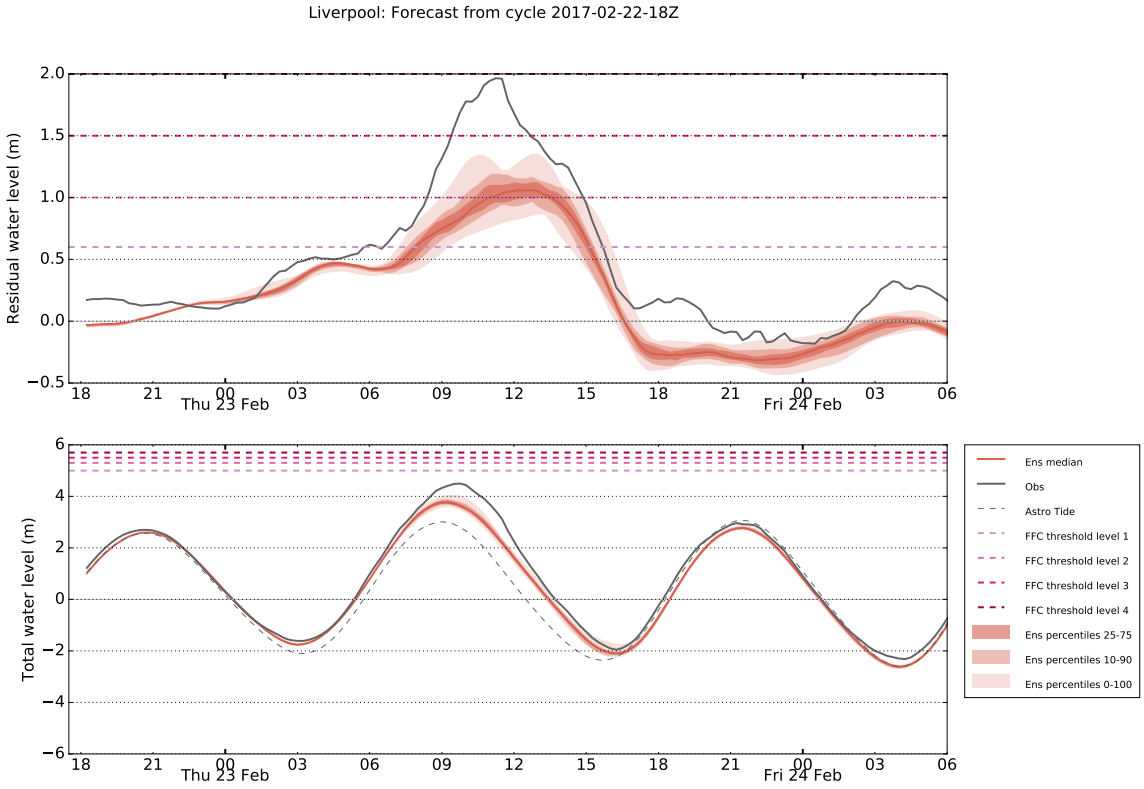

Figure 31: NEMO-Surge forecast for Liverpool from 2017-02-22 18Z run (storm Doris)

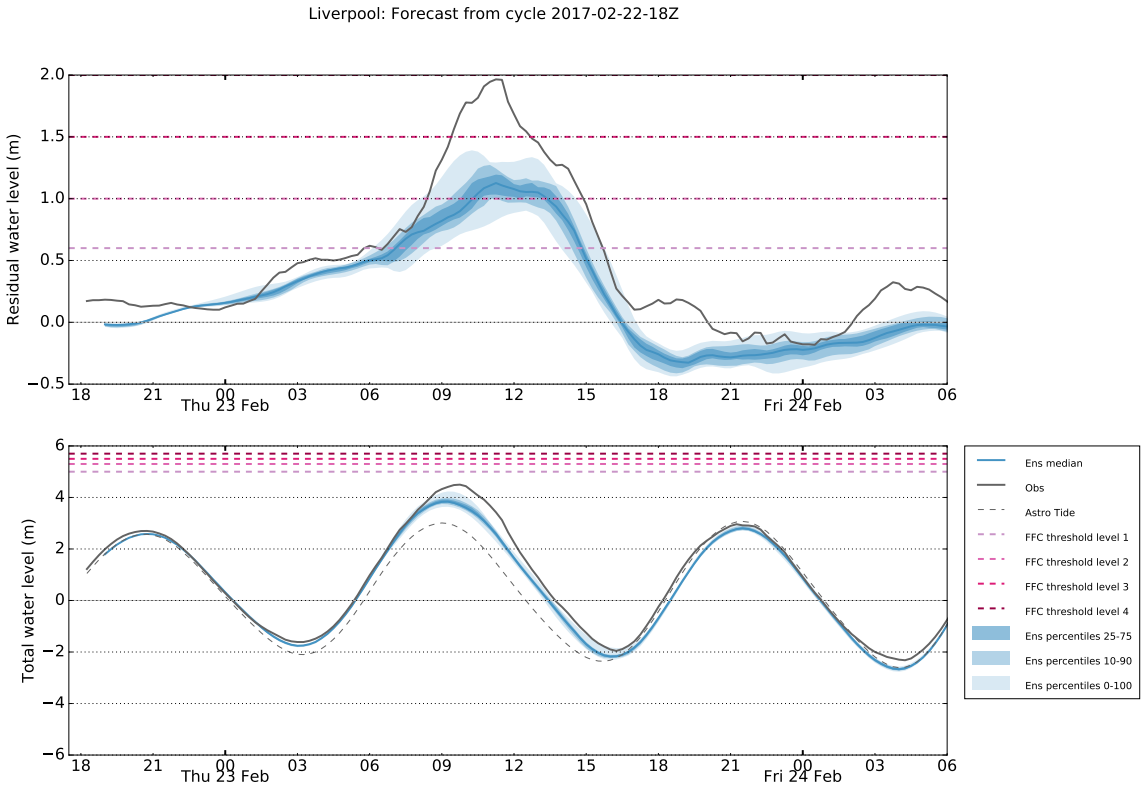

Figure 32: CS3X forecast for Liverpool from 2017-02-22 18Z run (storm Doris)

© Crown Copyright 2017

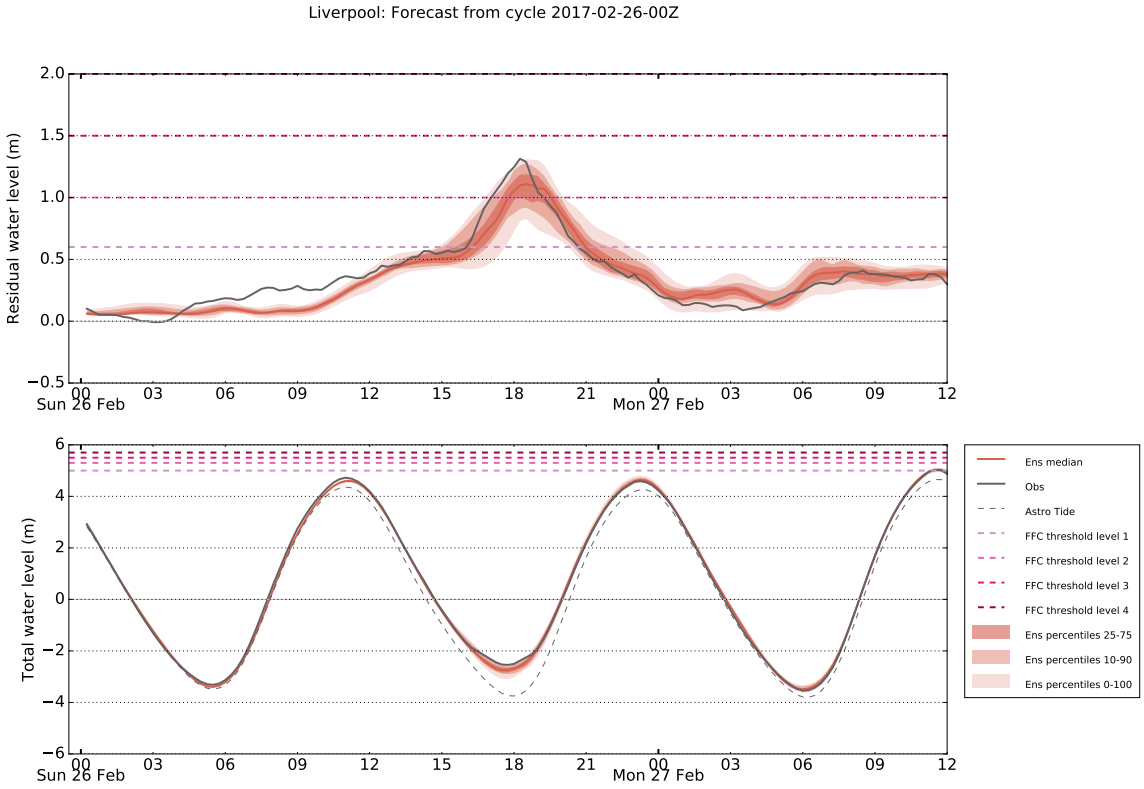

Figure 33: NEMO-Surge forecast for Liverpool from 2017-02-26 00Z run (storm Ewan)

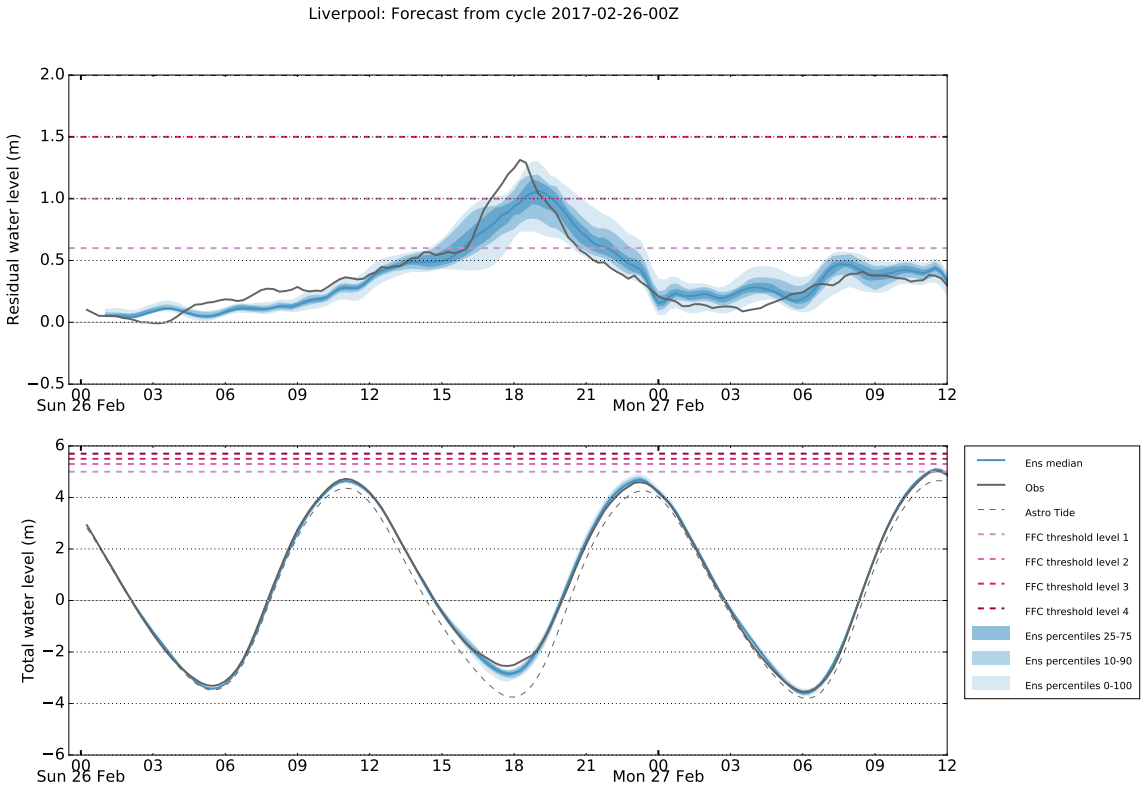

Figure 34: CS3X forecast for Liverpool from 2017-02-26 00Z run (storm Ewan)

© Crown Copyright 2017

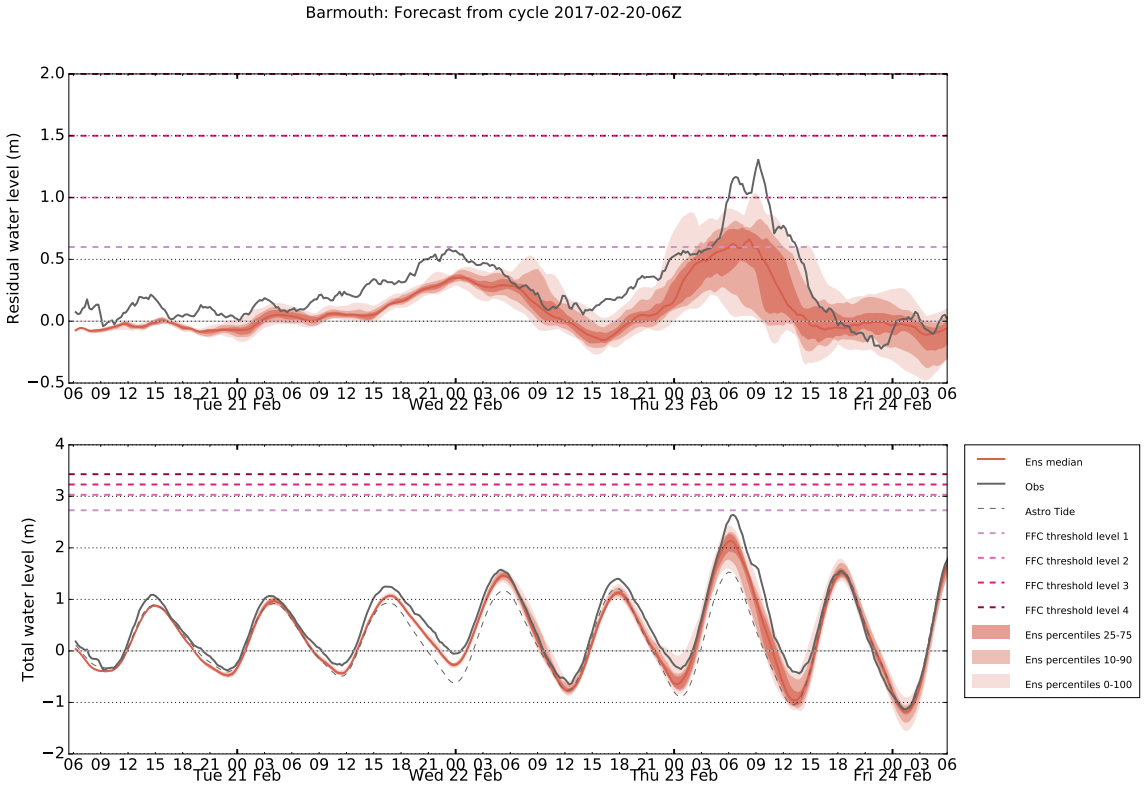

Figure 35: NEMO-Surge forecast for Barmouth from 2017-02-20 06Z run

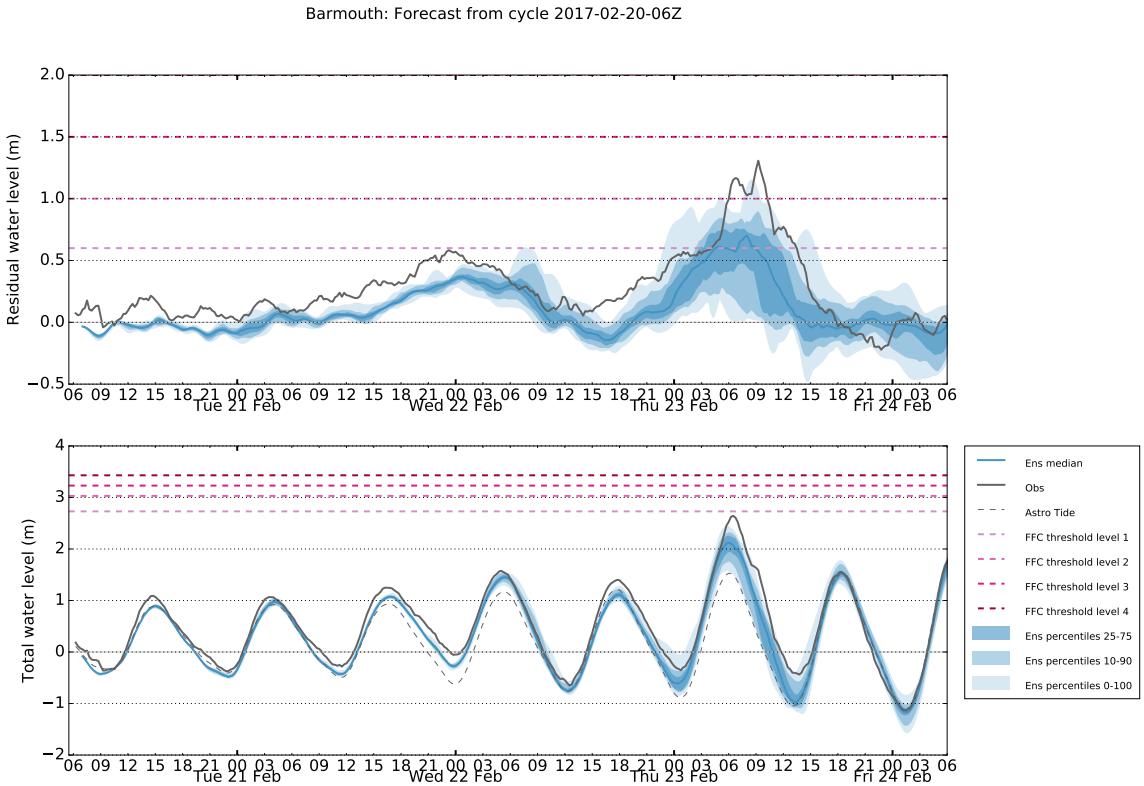

Figure 36: CS3X forecast for Barmouth from 2017-02-20 06Z run

© Crown Copyright 2017

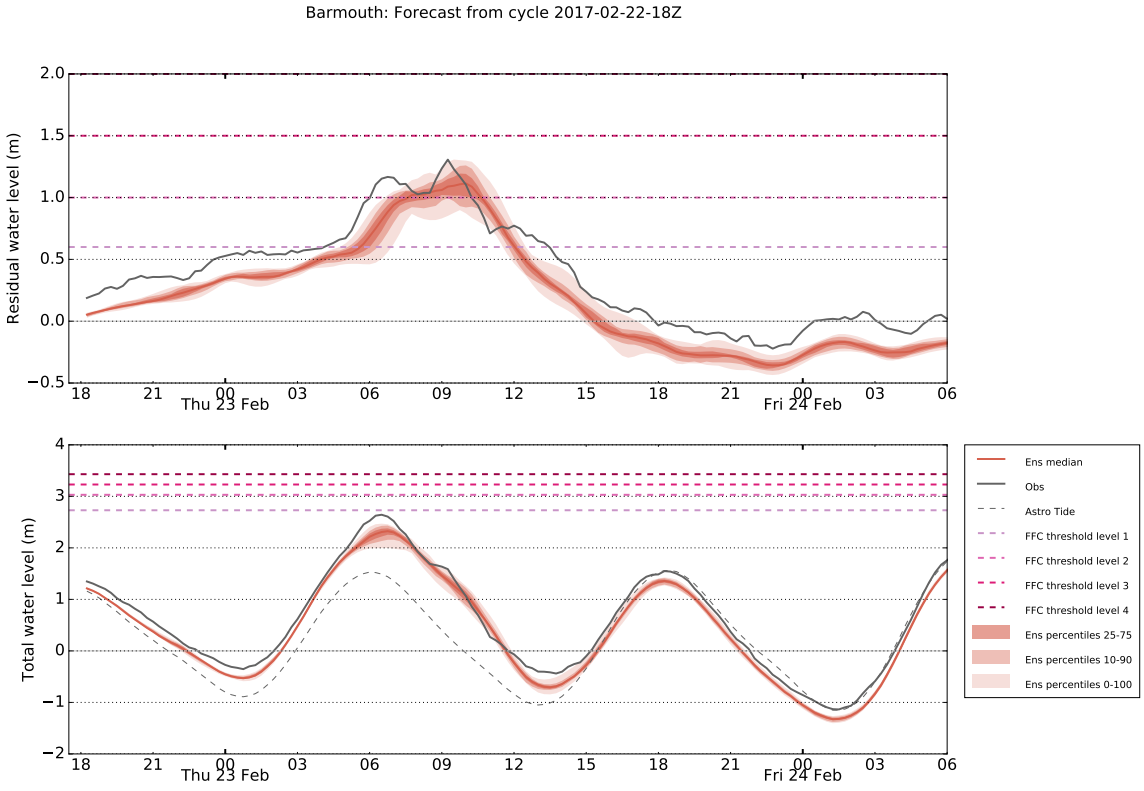

Figure 37: NEMO-Surge forecast for Barmouth from 2017-02-22 18Z run

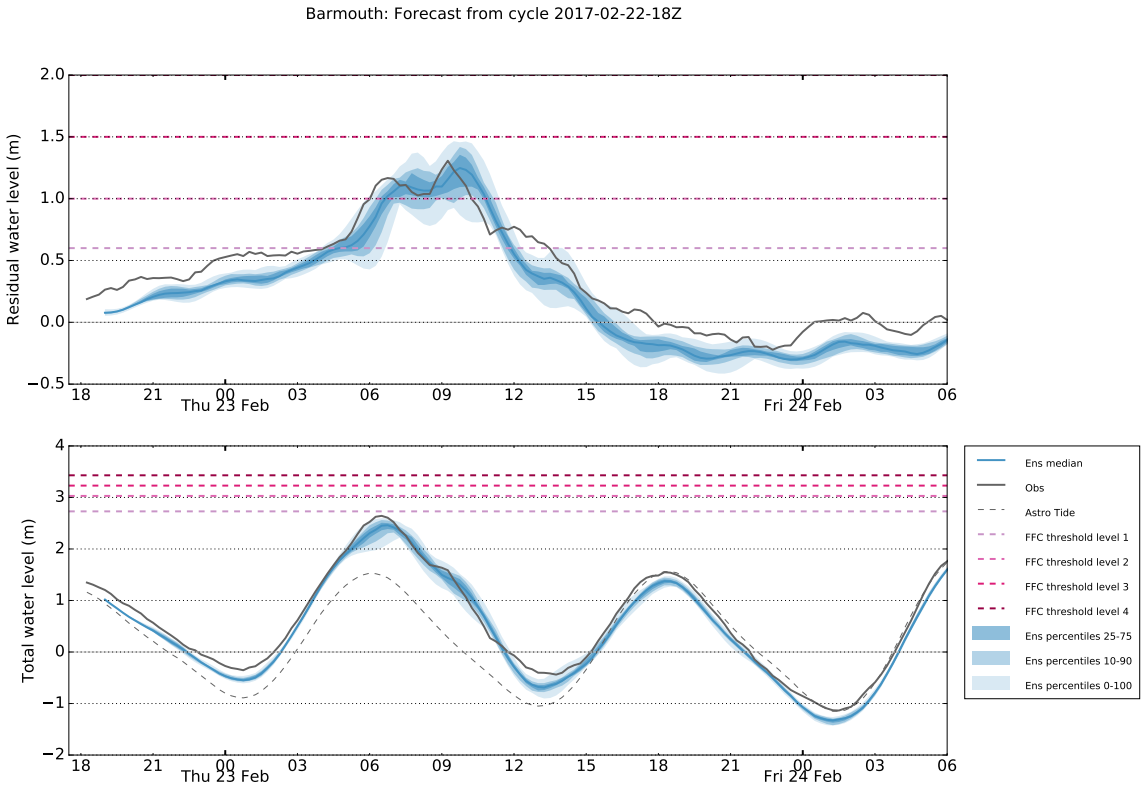

Figure 38: CS3X forecast for Barmouth from 2017-02-22 18Z run

© Crown Copyright 2017

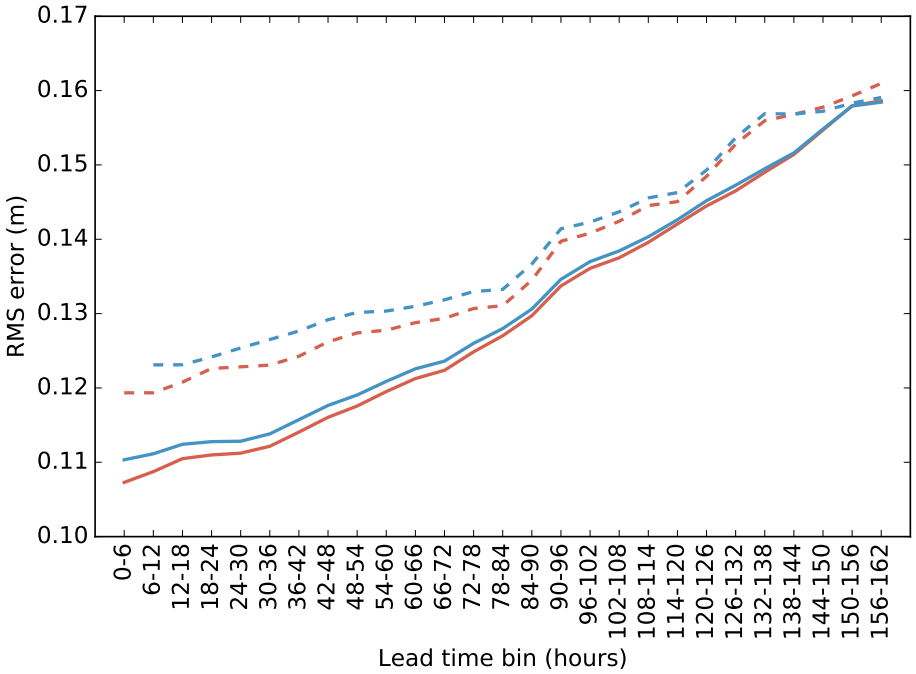

Figure 39: RMS error at Barmouth for residual (solid lines) and skew surge (dashed lines) for NEMO-Surge (red) and CS3X (blue)

6 Case studies: South coast

6.1 13–14 January 2017

The 13 January event described above in section 4 also generated a relatively large surge at the more eastern ports on the South coast such as Portsmouth and Newhaven. The residual peaked on the night of the 13th into the early hours of the 14th, coinciding with spring high tide at these locations.

Figures 40 & 41 (Newhaven) and 42 & 43 (Portsmouth), show that in this region the surge was well forecast by both models, with the observed residual falling well within the model ensemble. NEMO-Surge and CS3X are again very similar to each other in both cases. At Portsmouth the timing of the surge is slightly better in NEMO-Surge, leading to a better peak in the total water level.

© Crown Copyright 2017

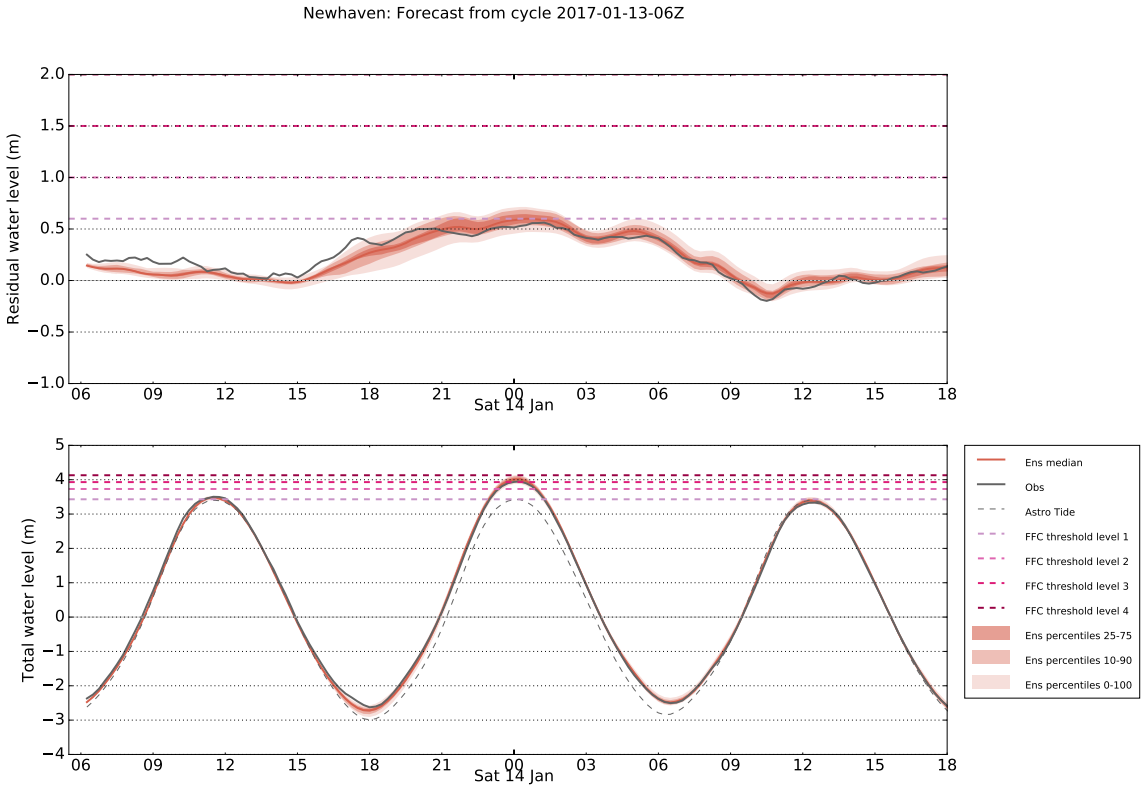

Figure 40: NEMO-Surge forecast for Newhaven from 2017-01-13 06Z run

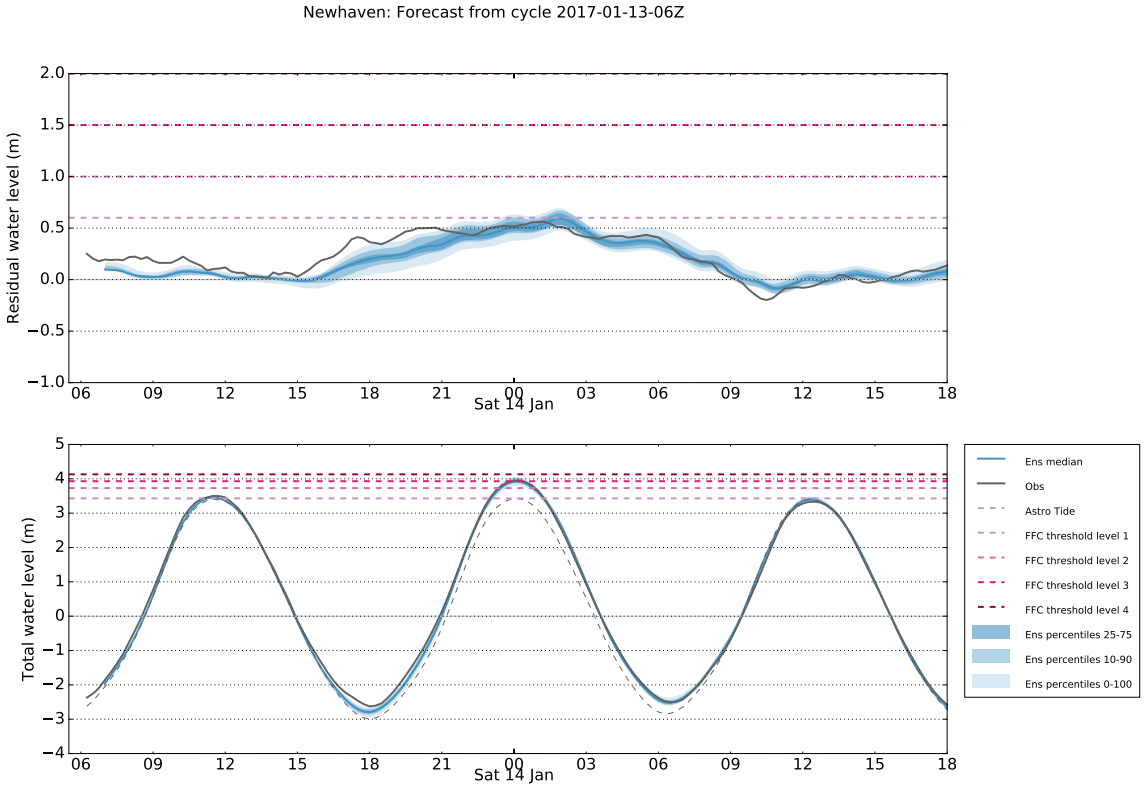

Figure 41: CS3X forecast for Newhaven from 2017-01-13 06Z run

© Crown Copyright 2017

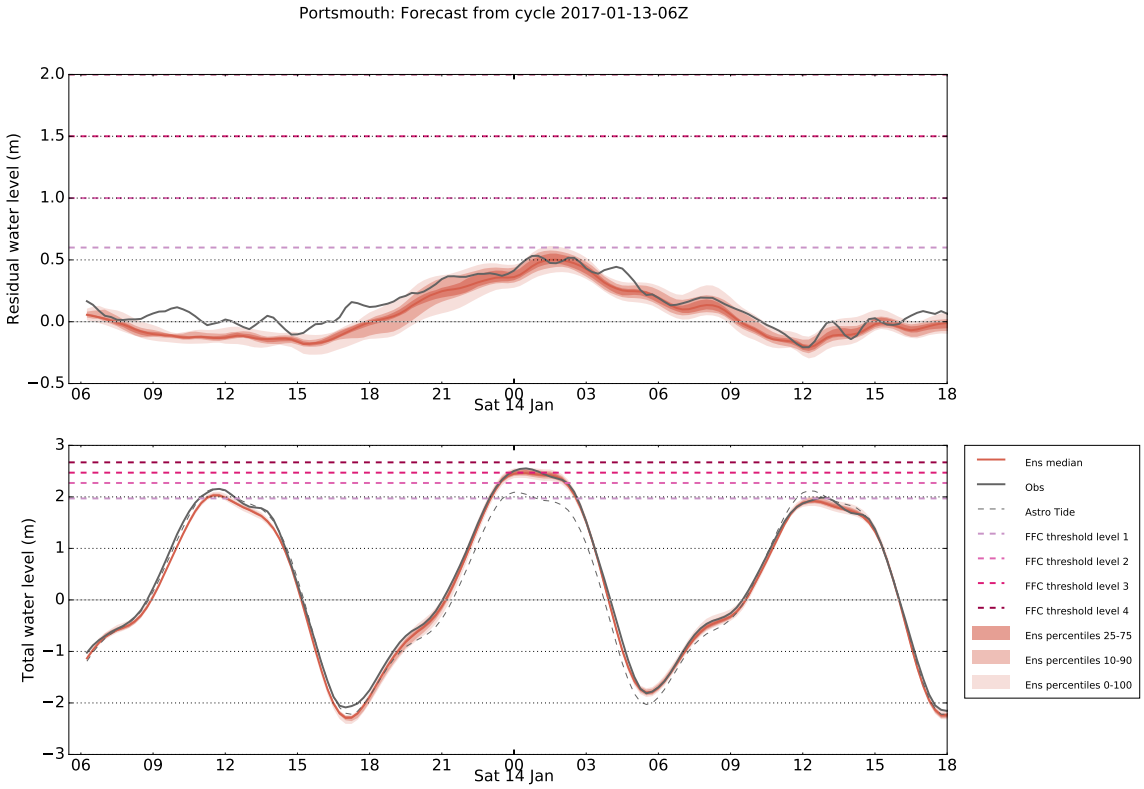

Figure 42: NEMO-Surge forecast for Portsmouth from 2017-01-13 06Z run

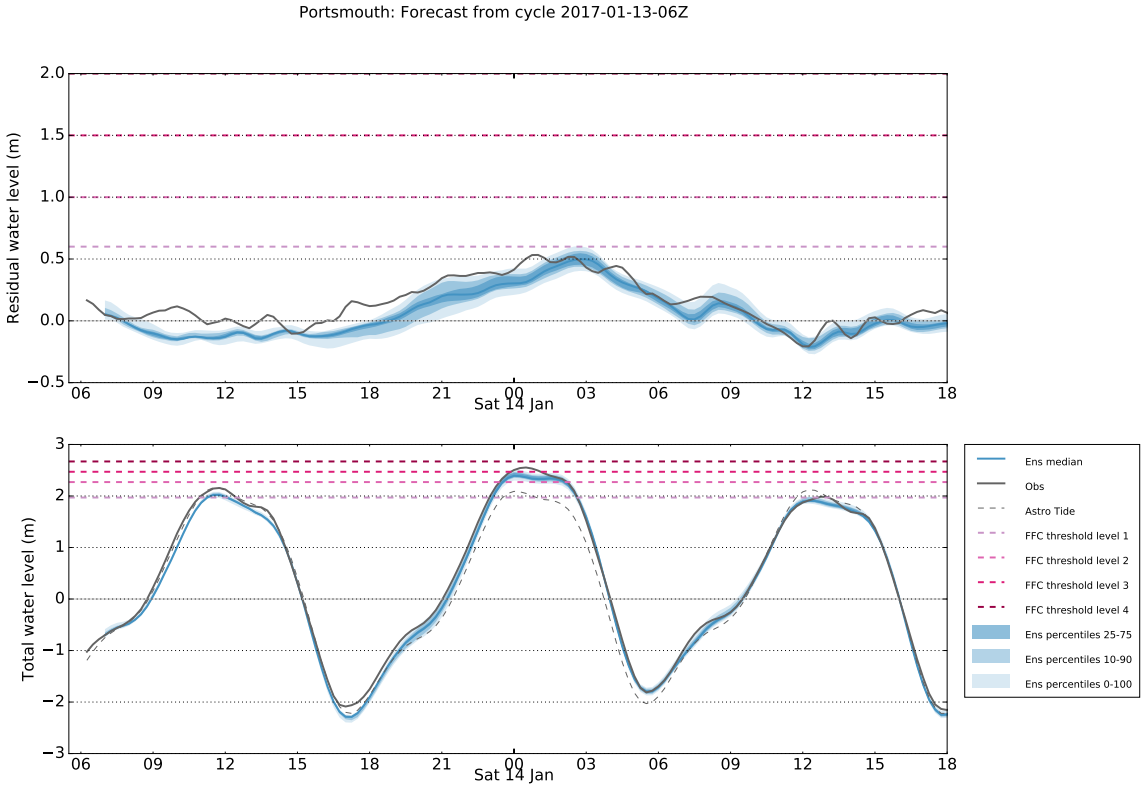

Figure 43: CS3X forecast for Portsmouth from 2017-01-13 06Z run

© Crown Copyright 2017

6.2 28–29 May 2017

A small surge was observed at some locations on the South cost at the end of May, outside the main storm season at the end of May. Figures 45 and 46 show an example for Bournemouth which had the largest signal, and we see that it was not picked up well by the model forecasts. Both NEMO-Surge and CS3X do show a general increase in residual surge level, though there is a general bias at this location with the residual being initially negative and then increasing to only around 0. This background bias is likely to be at least partly due to the issues already identified in section 3.1. NEMO-Surge and CS3X are indistinguishable from each other.

It is also noticeable that the model forecasts are smoother than the observed residual which has more high frequency features. The atmospheric charts from this period (Figure 44) show a complex system of fronts and it is possible that the surge was driven by high frequency gusty winds which are not resolved by the global atmospheric model that is used to force the surge models.

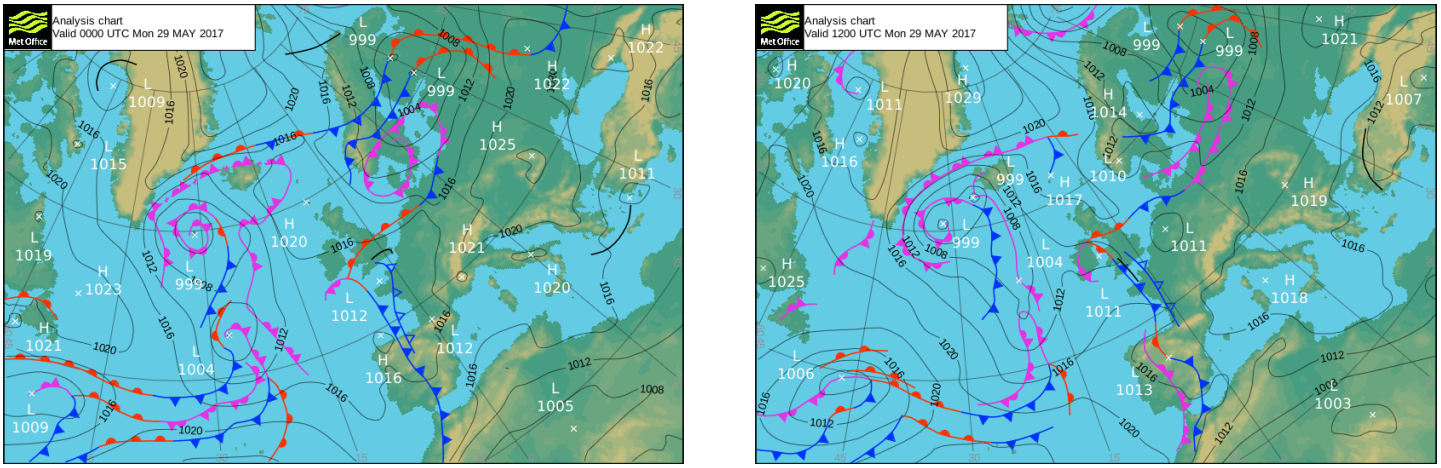

Figure 44: Analysis charts for 00Z (left) and 12Z (right) on 2017-05-29

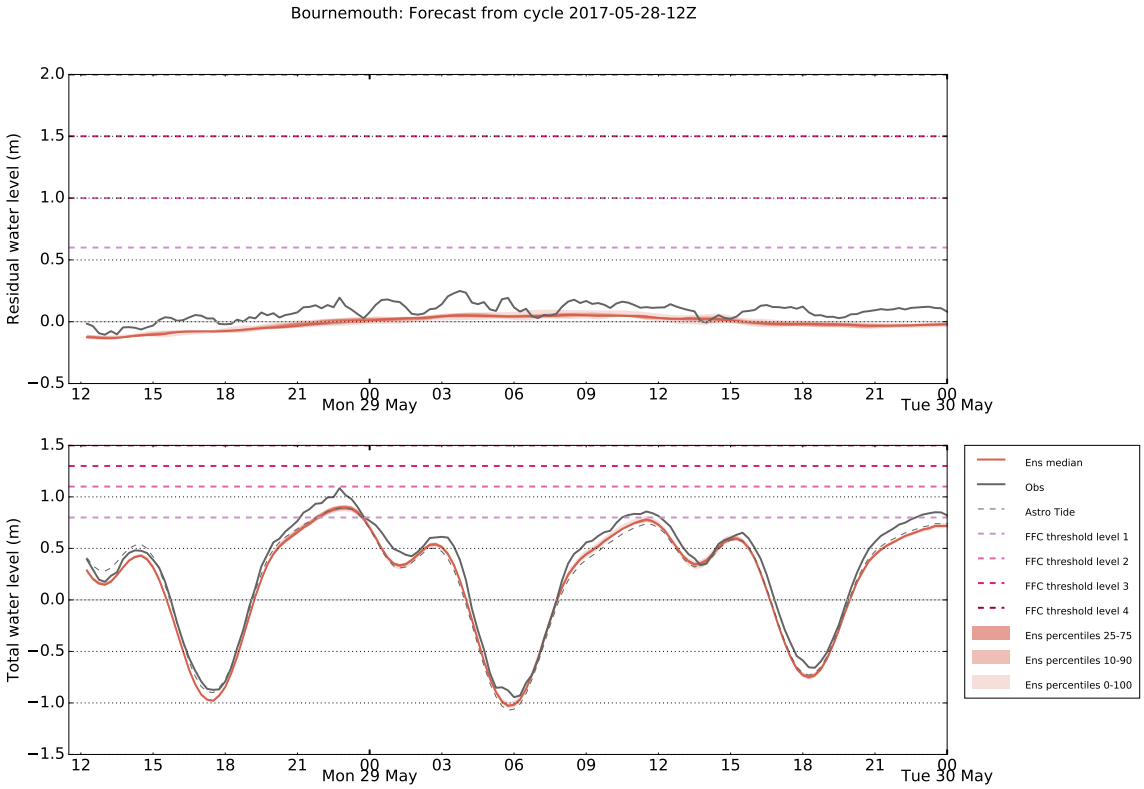

Figure 45: NEMO-Surge forecast for Bournemouth from 2017-05-28 12Z run

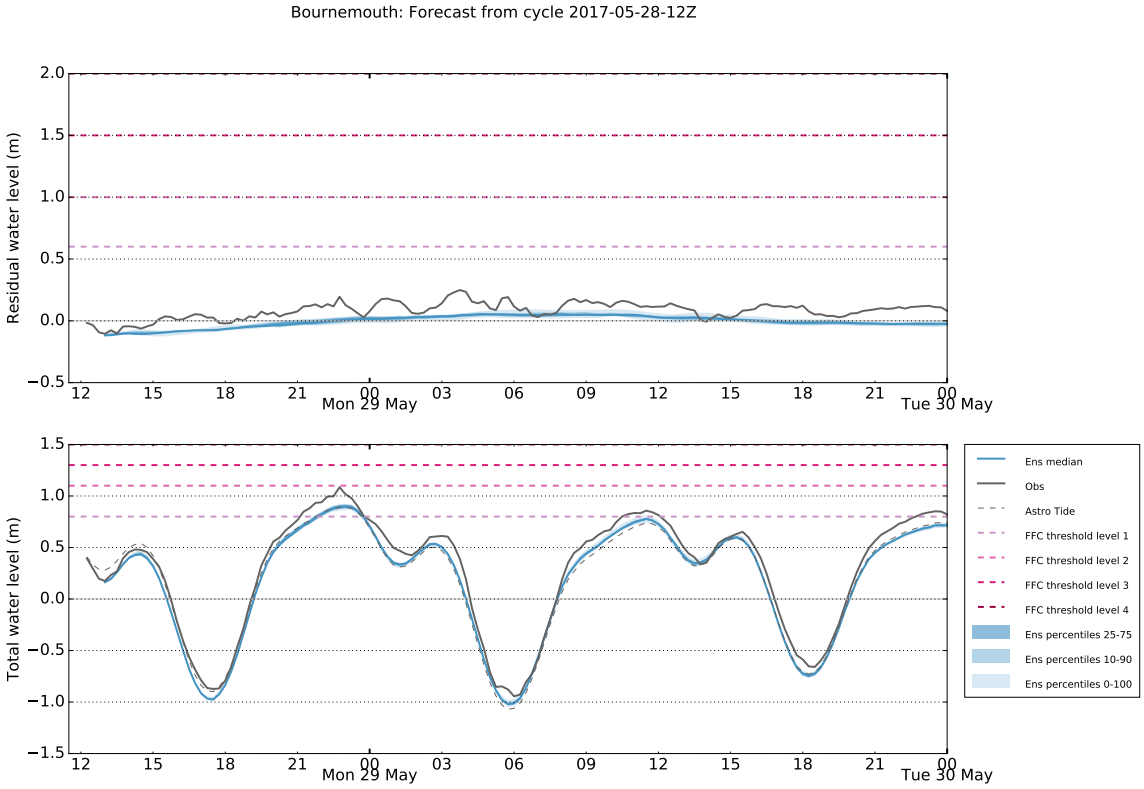

Figure 46: CS3X forecast for Bournemouth from 2017-05-28 12Z run

© Crown Copyright 2017

7 Overall statistics

Figures 47 and 48 show statistics, for the residual and skew surge respectively, computed by comparing NEMO-Surge and CS3X ensemble means against observations then taking an average over all of the observation locations.

As would be expected, the statistical performance for most metrics improves as the leadtime decreases. The ratio of standard deviation improves the most between 7 days and 3–4 days' leadtime and then levels out, indicating that the ensemble becomes focussed on the major signal in the surge at 3–4 days ahead. The RMS error on the other hand continues to improve past this point right up to the shortest leadtime, suggesting that the shorter leadtime forecasts are important for getting the finer details of surge height and timing correct.

The exception is the bias which in fact worsens with decreasing leadtime, although the values are small ranging from around -2 to -4 cm. Looking at individual ports rather than this overall average, most ports show the same behaviour though a small number do show improvement with decreasing leadtime (eg Workington shown in Figure 49, note change in scale).

This counter-intuitive behaviour may be to be due to the issues mentioned in Section 3.1 with certain long-period components being included in both the tide and residual, leading to an improving forecast seeming to move away from the 'truth'. It could also be due to a drift in the atmospheric model forcing. The important result for consideration here however is that NEMO-Surge behaves in the same way as CS3X.

For residual surge (Figure 47) there is virtually no difference between NEMO-Surge and CS3X for any of the metrics considered. The skew surge (Figure 48) shows slightly more difference in the bias and standard deviation ratio, but this is still small with only around 5 mm difference in the bias between the two models.

© Crown Copyright 2017

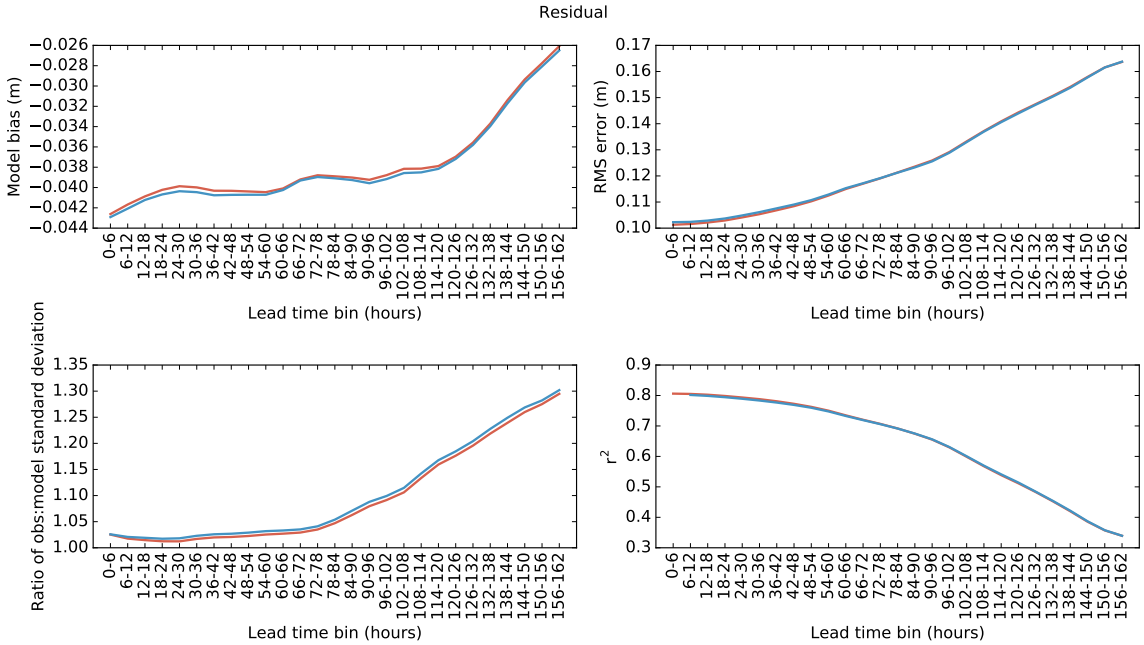

Figure 47: Statistical comparison for residual, averaged over all port locations, for NEMO-Surge (red) and CS3X (blue)

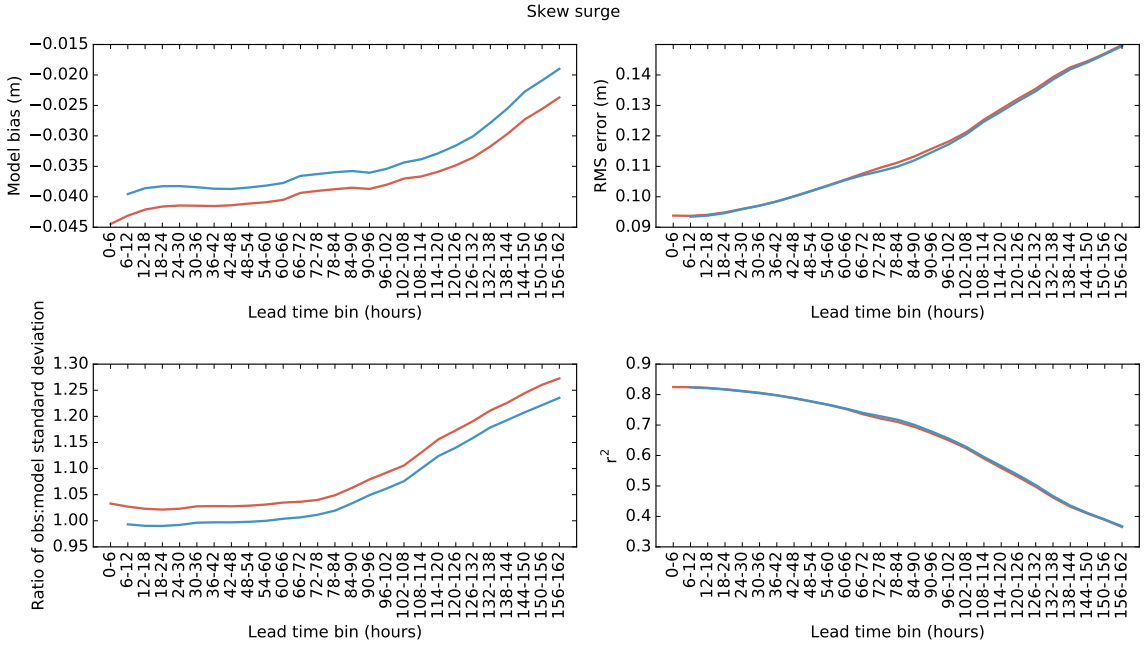

Figure 48: Statistical comparison for skew surge, averaged over all port locations, for NEMO-Surge (red) and CS3X (blue)

© Crown Copyright 2017

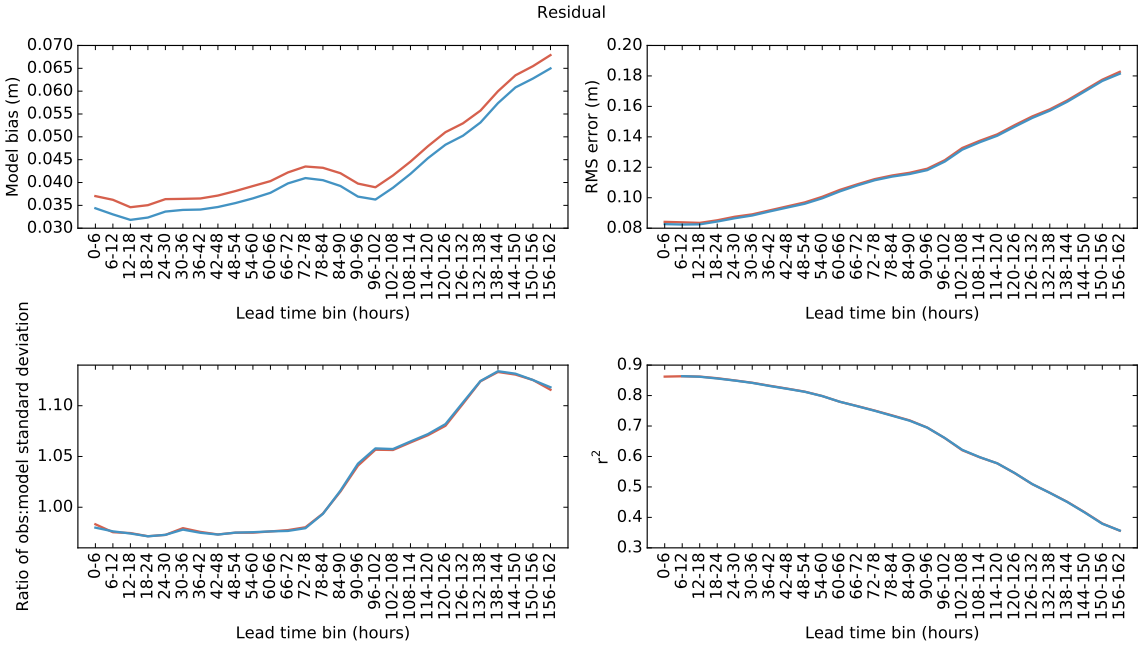

Figure 49: Statistical comparison for residual at Workington, for NEMO-Surge (red) and CS3X (blue)

8 Discussion and conclusions

NEMO-Surge was successfully introduced into the Met Office operational system in November 2016 and has been running stably since then. Overall statistical comparison shows very little difference between NEMO-Surge and CS3X for residual and skew surge. This is echoed in the case study results where, although there are sometimes small differences between the models most likely associated with variations in the model tide, the overall picture is very similar and they both display the same successes and limitations in forecast performance.

In the next stages of the project, NEMO-Surge will continue to run in parallel with CS3X while we work with customers to prepare them to receive trial data from NEMO-Surge for assessment in their downstream processing systems.

Although the case studies show good forecast performance in many cases, there were nevertheless events which were significantly over- (Sheerness, 13 January) or under- (Liverpool, storm Doris) predicted. It is clear therefore that in the longer term there is a need for model improvements to be made. Before this can be done, however, it will be necessary to address the underlying biases caused by the tide prediction data that were identified in Section 3.1, as well as further investigation of any spurious trends in the gauge observations such as that reported by Byrne et al. (2017) for the Bristol Channel area.

This will also require engagement from customers in order to understand how their present post-

© Crown Copyright 2017

processing systems are accounting for biases in the model (i.e. 'astro re-engineering' techniques), as well as supporting them to update these systems following any changes to improve the surge forecast products that they receive.

As well as the surge model itself, the atmospheric forecast used to force the model is also clearly important to get an accurate surge forecast. Shortly after the period considered in this report, in June 2017 the Met Office MOGREPS-G system underwent a significant upgrade including an increase in horizontal resolution from 33 km to 20 km. As this system is currently used as the forcing for the surge models, it is hoped that any improvements in the atmospheric forecast will in turn lead to improvements in the surge forecast for the 2017–18 storm season.

Additionally, work is planned to assess the feasibility of using the ECMWF global medium range forecast as input forcing to the surge model. This would potentially allow for a larger ensemble (51 members vs present 18+18) and longer forecast period (15 days vs 7 days).

In conclusion, the results from this initial forecast trial period have echoed the results of the hindcast runs analysed by O'Neill et al. (2016), with NEMO-Surge continuing to show very similar performance to CS3X. NEMO-Surge can therefore be seen as a viable replacement for CS3X for providing operational forecasts.

References

Bowler, N.E., Arribas, A., Mylne, K.R., Robertson, K.B., and Beare, S.E. (2008), The MOGREPS short-range ensemble prediction system, *Quarterly Journal of the Royal Meteorological Society*, *134*(632), 703–722, DOI: 10.1002/qj.234 .

Byrne, D., Robbins, G., Counsell, N., How, A., Saulter, A., O'Neill, C., and Pope, J. (2017), Improving sea level forecasting at Newport, *Internal project report*, Flood Forecasting Centre and Natural Resources Wales.

Flather, R., Williams, J., and Blackman, D. (2000), Causes of seasonal sea level variations and implications for surge prediction, *POL internal document 136*, Proudman Oceanographic Laboratory [now NOC].

Flather, R.A. (1994), A storm surge prediction model for the northern Bay of Bengal with application to the cyclone disaster in April 1991, *Journal of Physical Oceanography*, *24*(1), 172–190, DOI: 10.1175/1520-0485(1994)024<0172:ASSPMF>2.0.CO;2 .

Flowerdew, J., Horsburgh, K., Wilson, C., and Mylne, K. (2010), Development and evaluation of an ensemble forecasting system for coastal storm surges, *Quarterly Journal of the Royal Meteorological Society*, *136*(651), 1444–1456, DOI: 10.1002/qj.648 .

© Crown Copyright 2017

Furner, R., Williams, J., Horsburgh, K., and Saulter, A. (2016), NEMO-surge: Setting up an accurate tidal model, *Weather science technical report 610*, Met Office. URL: http://www.metoffice.gov.uk/binaries/content/assets/mohippo/pdf/6/m/frtr_610_2016p.pdf, [accessed Aug 2017].

Haigh, I.D., Wadey, M.P., Wahl, T., Ozsoy, O., Nicholls, R.J., Brown, J.M., Horsburgh, K., and Gouldby, B. (2016), Spatial and temporal analysis of extreme sea level and storm surge events around the coastline of the UK, *Scientific Data*, *3*(160107), DOI: 10.1038/sdata.2016.107 .

National Tidal and Sea Level Facility (2015), Annual report for 2015 for the UK National Tide Gauge Network, *Tech. rep.*, BODC. URL: https://www.bodc.ac.uk/data/hosted_data_systems/sea_level/uk_tide_gauge_network/reports/documents/2015annualreport.pdf, [accessed Aug 2017].

O'Neill, C., Saulter, A., Williams, J., and Horsburgh, K. (2016), NEMO-surge: Application of atmospheric forcing and surge evaluation, *Weather science technical report 619*, Met Office. URL: http://www.metoffice.gov.uk/binaries/content/assets/mohippo/pdf/library/frtr_619_2016p.pdf, [accessed Aug 2017].

UK Storm Centre (2017). URL: http://www.metoffice.gov.uk/barometer/uk-storm-centre/uk-storm-centre, [accessed Aug 2017].

© Crown Copyright 2017

Met Office
FitzRoy Road, Exeter
Devon, EX1 3PB
UK

Tel: 0370 900 0100
Fax: 0370 900 5050
enquiries@metoffice.gov.uk
www.metoffice.gov.uk

Produced with the Met Office LaTeX template v2.0.0